

ON THE OPTIMAL GEOGRAPHICAL ORGANISATION OF
PUBLIC SERVICES WITH CENTRAL FACILITIES

ROBERT LINDSAY HODGART

DOCTOR OF PHILOSOPHY
UNIVERSITY OF EDINBURGH

1981



This thesis has been composed by me and is based on my own work.

July 1981

Some material included in the thesis (mainly in Chapters 5 and 6) has already been published and a copy of the paper is enclosed at the end.

Abstract of Thesis

The thesis is concerned with concepts and methods which are relevant to locating facilities to serve surrounding catchment areas (for instance public libraries and swimming pools). When demand does not depend on the location of centres (i.e. it is inelastic), locations can be found which minimize the travel costs of users over a system of Thiessen catchments by existing heuristic algorithms. In the thesis a new algorithm is developed for locating centres on a plane to maximize consumers' level of use, when demand is elastic.

The question of the optimal number of centres is also explored by developing a conceptual framework which, under various assumptions, shows the relation between the number of centres and the overall level of use and travel costs on an isotropic plane. Using this framework, it is possible to demonstrate how the optimal number of centres depends on the elasticity of demand and on the extent to which supply can be dispersed in smaller units without affecting costs.

The thesis shows that centres can be located on a plane to maximize use through a method of spatial search based on the partial derivatives of the objective function. A location/allocation type of algorithm (called LOCHWISP) employing this principle is developed successfully both for Thiessen and overlapping catchment areas, the latter being defined by using a spatial interaction model. The algorithm is heuristic and produces convergent results.

LOCHWISP is then used to examine various sites in Edinburgh which were considered as alternative locations for the Commonwealth Pool and to assess a number of sites proposed for district swimming pools. It is suggested that LOCHWISP would have been a useful aid in the

process of decision making, especially in providing better estimates of how the new pools would affect the level of use at the older ones.

ACKNOWLEDGEMENTS

Since its inception a long time ago, this thesis has been done on a part-time basis. As it had to be done largely in the summer months it has not been easy to complete and I am therefore more than grateful to all those, too numerous to mention by name, who have helped in different ways to bring it to a conclusion.

My first thanks are to my supervisor, Prof. J.T. Coppock. His comments have stimulated several new lines of thought; his editorial advice has led to many improvements in the text; and by helping to arrange leave for me in the winter of 1976-77 he enabled me to make much progress with the first draft. Not least, his enthusiasm is always infectious. To Peter Gould, who first interested me in the problem, and to Isobel Robertson, who provided valuable advice in the early stages, I also owe special thanks.

The work really took off in the summer of 1975 and I am particularly grateful to several people in Edinburgh who helped to push the crate onto the runway then and pointed the pilot in the right direction. Bob Donaldson, then doing a Ph.D. in Chemical Engineering, celebrated EMAS long and patiently enough in my ears to gain a convert. John Martin in Mathematics provided some vital clues by helping me to relearn calculus. He also gave me some idea of its potential by showing me how to calculate mean travel distance from a circular catchment area. At a later date he repeated this calculation for a hexagonal catchment and checked the partial derivatives I had worked out for various functions. Ken McKinnon, also of Mathematics, provided much good advice on methods of optimization and other matters mathematical. Any mathematical errors in the text are, of course, mine.

Without the advisory staff of ERCC, particularly Morton Ogilvie, Malcolm Brown and Bill Watson, it would have taken me much longer to get NORLOC and CONVERTS to work at a crucial point in the summer of 1975. Jack Hotson's assistance with CAMGRID enabled me to draw maps of numerous social and demographic variables for Edinburgh quickly and easily and gain a useful background picture of the city's social geography. The operating system at Edinburgh, EMAS, is a great boon to all users and I would like to let its authors, whoever they are, know that it is appreciated.

I have benefitted greatly from contacts with the Tourist and Recreation Research Unit in the Geography Dept. The presence in TRRU in the summers of 1976 and 1977 of Prof. Gordon Ewing was particularly stimulating. Brian Duffield detected an inconsistency in a paper I was about to publish, while Mike Owen provided encouragement and help on the applied side. Jonathan Long kindly lent me several useful papers. Chuck Chulvick of PLU (then with PDMS) provided the 1971 census data from his SPSS files. I really owe the whole unit a general word of thanks.

I am also fortunate in often being able to get immediate help from several other people who frequent the computing room in Geography. Steve Dowers (also of TRRU) has saved me valuable time on several occasions through his wide knowledge and quick diagnosis. I feel a similar debt to Craig Stott. At one time or another Roger Musson, Vicky Eachus and Tom Waugh have also provided clues or solutions regarding EMAS.

I would also like to give a special word of thanks to one of my postgraduate students, John McCalden, for his advice on various computing matters and for his resilience and good humour in tackling all kinds of problems. In particular, John used GIMMS to draw a

map of the distribution of population in Edinburgh (used as Figure 3.5) which was much more legible than the map I had drawn by line printer. In addition, he provided the core of the subroutine which enabled the searching path traced by LOCHWISP to be mapped by graph plotter.

Charles ReVelle, Gerry Rushton, Michael Goodchild and Barry Boots sent me many valuable papers and Michael Goodchild also provided advice on LAP. Bryan Massam's pioneering work on several aspects of the problem has furnished valuable building material, not least of a bibliographic kind. Laurence Ostresh's Ph.D. thesis was an important stimulus. Two months of my leave in 1977 were spent in Poland and I would like to record my gratitude to many people there who took time to discuss their research work with me. Perhaps I can simply mention Professor Andrzej Hopfer of ART in Olsztyn as a representative of all the others, since he has helped in so many ways.

I am also grateful to Arne Foldvik of the Institute of Linguistics at University of Trondheim, who provided the main information on Norwegian schools used in Chapter 2. For most of the diagrams I am indebted to the draughtsmanship of Ray Harris and Mhorag Innes and the efficiency of Lorna Gowans in the dark room. Mrs. E. Miller typed the mathematical chapters patiently and very competently and Jan Hollingdale typed the others with great speed and accuracy.

Between the final typing of the text itself and the completion of diagrams and other material, I had to spend some months in hospital following an accident. Recovery went almost hand in hand with completing the work. Everyone in the Geography Department helped in some way during my recovery - the support and encouragement has been

simply wonderful. Sandy Crosbie, Head of the Department, and his wife Nicola have helped me in a host of ways. Professor Coppock has made several things easier for me, though still convalescing himself and I am also particularly grateful to Dave Lennie and Douglas Hunter for taking much of the xeroxing off my hands.

I must thank all the citizens of the South Side of Edinburgh whose good humour makes the area such a refreshing place to work. My parents have been a great support and inspiration all this time. I can never thank them adequately.

Table of Contents

	Page
Declaration	i
Abstract	ii
Acknowledgements	iv
Table of Contents	viii
List of Figures	xii
List of Tables	xvi
 <u>SECTION I</u>	
<u>The economic and political background of the problem of location</u>	
<u>Chapter 1</u>	2
Introduction	2
Some general characteristics of private and public services	5
The spatial organisation of public services	9
(a) Lack of competition	9
(b) Accessibility as an external benefit	15
(c) Influence of the political process	19
Formal structure of the problem	23
The role of access	26
Optimizing versus satisficing models	33
Focus and organisation of the thesis	35
The time dimension and other assumptions	38
 <u>SECTION II</u>	
<u>A framework for optimizing the number of facilities</u>	
<u>Chapter 2</u>	41
Introduction	41
Supply strategies	41
Economies of scale	42
Diseconomies of scale	52
Conclusions on effects of scale	54
 <u>Chapter 3</u>	
<u>The benefit to users of better access to facilities</u>	
Aim	57
Size of catchment areas as m increases	58
Aggregate and mean travel within one catchment area, given inelastic demand	62
Aggregate and mean travel for the whole system with inelastic demand	66
Aggregate travel in Edinburgh	68

	Page
Elastic demand	72
More flexible types of demand cone	85
Elastic demand with overlapping catchment areas	96
Net benefits to users	103
Equity	117
Covering goals	117
Conclusion	120
<u>Chapter 4</u> <u>The optimal number of centres</u>	
Introduction	122
Example A: inelastic demand	122
Example B: economies of scale	128
Example C: elastic demand	130
Example D: private suppliers	138
Example E: a spatially flexible supply of given amount	139
Conclusion from examples	142
<u>SECTION III</u> <u>Methods of locating centres</u>	
<u>Chapter 5</u> <u>Exact methods of solving the location problem in a network</u>	
Introduction	146
Forms of the location problem	146
Geometric frameworks	147
Combinatorial programming	151
Linear programming	155
Extensions of the linear programming formulation	157
Models for maximizing social benefit	159
Conclusion	161
<u>Chapter 6</u> <u>Methods of searching space on a plane</u>	
Introduction	163
Surface topography of different objectives	163
Locating one centre on a plane	170
Space-searching heuristics for m facilities	175
The Törnqvist algorithm	177
Comparison of Törnqvist and Cooper algorithms	179
Experiment with searching of interstices	182

	Page
<u>Chapter 7</u> <u>The LOCHWISP algorithm</u>	
Introduction	187
Experience with NORLOC and LAP	189
The structure of the LOCHWISP algorithm	193
Range of problems covered by LOCHWISP	198
Testing LOCHWISP	200
Searching properties of LOCHWISP	
(a) The location of one centre	208
(b) The location of several centres	213
(c) Searching with overlapping catchments	221
Conclusion	231
 <u>SECTION IV</u> <u>The location of swimming pools in Edinburgh</u>	
<u>Chapter 8</u> <u>The location of the Commonwealth Pool</u>	
Introduction	234
The location of pools in Edinburgh	235
History of swimming baths in Edinburgh	237
The decision to locate the Commonwealth Pool	238
Analysis of potential locations by LOCHWISP	245
Effect of the Commonwealth Pool on catchments and use of older pools	
(a) Predicted impact	255
(b) Actual impact of the new pool	263
Conclusion	268
 <u>Chapter 9</u> <u>The planning and location of district pools</u>	
Introduction	270
History of the main decisions regarding district pools	271
Explanation of priorities	274
Modification of population base	280
Strategy of locational analysis	282
Analysis of Wester Hailes as the site for an additional pool	283
Plans relating to the proposed pool at Comiston	297
Choice of two locations for district pools	300
Conclusion	304

		Page
<u>SECTION V</u>	<u>Conclusion</u>	
<u>Chapter 10</u>	<u>Conclusion</u>	310
Bibliography		321
<u>Appendix 1</u>	Main computer programs used in the work	329
<u>Appendix 2</u>	Location of one centre on a plane to maximize use	331
<u>Appendix 3</u>	Location of one centre of maximize NBU on a plane	333
<u>Appendix 4</u>	The LOCHWISP algorithm	336
<u>Appendix 5</u>	The effective extent of demand cones and catchment areas	344
<u>Appendix 6</u>	List of mathematical symbols	346

List of Figures

		Page
Figure 2.1	Assumed relationship between: (a) size of facility and cost of supplying one facility (b) number of facilities (m) and total cost of supply (SC)	56
Figure 3.1	Hexagonal catchment area with inscribed and circumscribed circles	60
Figure 3.2	Landscape in which the number of central points can have any integer value	63
Figure 3.3	Unwinding a ring within a circular catchment to form a rectangular strip	65
Figure 3.4	Relation between aggregate travel and m on an isotropic plane when demand is inelastic	69
Figure 3.5	The distribution of population in Edinburgh by grid squares of 500 m in 1971	71
Figure 3.6	Relation between aggregate travel and m for Edinburgh with centres located by NORLOC	73
Figure 3.7	Demand cones defined by inverse powers of distance	76
Figure 3.8	Relation between volume of use and m on an isotropic plane for various values of b (user response curve)	84
Figure 3.9	Comparison of distance decay curves based on inverse powers and on negative exponentials	86
Figure 3.10	Family of user response curves with decline in use described by a negative exponential function - absolute values	90
Figure 3.11	Family of user response curves with decline in use described by a negative exponential - relative to values with $m = 1$	93
Figure 3.12	Effect of overlapping catchments on user response curves when demand is strongly elastic ($b = 1$)	97
Figure 3.13	Effect of overlapping catchments on user response curves when demand is mildly elastic ($b = .25$)	98
Figure 3.14	Relation between m and NBU, BU, AT and MT, expressed as percentages of values for $m = 1$, on an isotropic plane with $b = .25$ and $v = 10$.	106

List of Figures (contd.)

	Page
Figure 3.15	Relation between m and NBU, BU, AT and MT, expressed as percentages of values for $m = 1$, on an isotropic plane with $b = .50$ and $v = 10$. 107
Figure 3.16	Relation between m and NBU, BU, AT and MT, expressed as percentages of values for $m = 1$, on an isotropic plane with $b = .75$ and $v = 10$. 108
Figure 3.17	Relation between m and NBU, BU, AT and MT, expressed as percentages of values for $m = 1$, on an isotropic plane with $b = 1.0$ and $v = 10$. 109
Figure 3.18	Relation between m and NBU, BU, AT and MT, expressed as percentages of values for $m = 1$, on an isotropic plane with $b = .25$ and $v = 5$. 112
Figure 3.19	Relation between m and NBU, BU, AT and MT, expressed as percentages of values for $m = 1$, on an isotropic plane with $b = .50$ and $v = 5$. 113
Figure 3.20	Relation between m and NBU, BU, AT and MT, expressed as percentages of values for $m = 1$, on an isotropic plane with $b = .75$ and $v = 5$ 114
Figure 3.21	Relation between m and NBU, BU, AT and MT, expressed as percentages of values for $m = 1$, on an isotropic plane with $b = 1.0$ and $v = 5$. 115
Figure 3.22	Comparison between a constant radius of cover, q , and the radius of catchment areas which changes as m increases 118
Figure 3.23	Relation between m and population covered 119
Figure 4.1	Relation between m , savings in travel (TS) and cost of supply (SC) 125
Figure 4.2	Relation between m , savings in travel (TS) and cost of supply (SC) illustrating economies (SC^E) and diseconomies (SC^D) of scale. 129
Figure 4.3	Relation between m , net benefit to users (NBU) and cost of supply (SC) for selected values of b 131
Figure 4.4	Relation between m , net benefit to users (NBU) and cost of supply (SC) when supply is spatially flexible (Example E) 140

List of Figures (contd.)

	Page
Figure 6.1 Distribution of population in hypothetical example	165
Figure 6.2 Aggregate travel at each village	165
Figure 6.3 Level of use of a facility sited at each village when $b = 1.0$	167
Figure 6.4 Maximum distance travelled to a facility sited at each village	167
Figure 6.5 Optimal locations derived from various objectives	169
Figure 6.6 Population covered from a facility sited at each village when $s = 3$	169
Figure 6.7 Search restricted by location of existing facilities	183
Figure 6.8 (a) Locations of a given set of points (b) Construction of a framework of triangles and definition of interstices	184
Figure 7.1 Structure of the LOCHWISP program	194
Figure 7.2 The family of locational problems which LOCHWISP can solve	199
Figure 7.3 Results of locating one centre in Edinburgh according to various objectives, using different starting locations	210
Figure 7.4 Lines of search for location of 7 centres in Edinburgh to minimize travel cost	214
Figure 7.5 Lines of search for location of 7 centres in Edinburgh to maximize use with $b = 0.500$	216
Figure 7.6 Lines of search for location of 7 centres in Edinburgh to maximize NBU with $v = 10.0$ and $b = 0.500$	218
Figure 7.7 Lines of search for location of 7 centres in Edinburgh to maximize NBU with $v = 5.0$ and $b = 0.250$	219
Figure 7.8 Lines of search for location of 7 centres in Edinburgh to maximize use with $b = 0.125$	226

List of Figures (contd.)

		Page
Figure 7.9	Lines of search for location of 7 centres in Edinburgh to maximize use with $b = 0.125$	227
Figure 7.10	Lines of search for location of 7 centres in Edinburgh to maximize use with $b = 0.250$	228
Figure 8.1	Reference map of Edinburgh	
	A fold out map of Edinburgh is provided inside the back cover and the reader will probably find this particularly helpful for Chapters 7, 8 and 9.	

List of Tables

	Page
Table 2.1 Summary chart showing existence of economies (or diseconomies) of scale	47
Table 2.2 Cost curve studies of scale economies	48
Table 3.1 Demand response values expressed in relative terms for different values of a_1	95
Table 3.2 Demand response values expressed as a percentage of total demand available	101
Table 4.1 Response of TRIPS, MT and NBU to increase in m from $m = 1$ to $m = 30$ with $b = 0.25$ and $v = 10$	135
Table 4.2 Optimal values of m in a marginal framework for given thresholds of NBU	136
Table 7.1 Comparison of solutions by LAP, LOCHWISP and NORLOC to a standard problem with seven centres	202
Table 7.2 Characteristics of search for location of one centre in Edinburgh by LOCHWISP	211
Table 8.1 Indices of accessibility for all pools with a new pool at various sites, expressed relative to the values for Holyrood Park Road, assuming Thiessen catchments	249
Table 8.2 Cost of travel and level of use for all pools with a new pool at various sites, expressed relative to the values for HPR, assuming overlapping catchments	251
Table 8.3 Cost of travel and level of use for all pools with a new pool at various sites, expressed relative to the values for HPR, using overlapping catchments and taking attractiveness into account	254
Table 8.4 Effect of locating 'central pool' at various sites on older pools, assuming Thiessen catchments	256
Table 8.5 Percentage reduction in size of catchments of older pools with Commonwealth Pool located at various sites, assuming overlapping catchments	258
Table 8.6 Percentage reduction in predicted level of use of older pools with Commonwealth Pool located at various sites, assuming overlapping catchments	260
Table 8.7 Percentage reduction in predicted level of use of older pools with Commonwealth Pool located at various sites, assuming overlapping catchments and taking attraction into account	261

List of Tables (contd.)

		Page
Table 8.8	Admissions to pools for the years 1968-69 and 1971-72	264
Table 9.1	Indices of accessibility with a new district pool at specific sites, relative to the values for W. Hailes, assuming Thiessen catchments	284
Table 9.2	Indices of accessibility with a new pool at specific sites relative to values for W. Hailes, using overlapping catchments and attractiveness	286
Table 9.3	Facility index values for various sites, given as percentage of values at W. Hailes (without considering attractiveness)	289
Table 9.4	Facility index values for various sites, given as percentage of values at W. Hailes with attractiveness considered	291
Table 9.5	Predicted level of use of older pools with new pool at various sites as a percentage of the original value, using overlapping catchments and attractiveness	292
Table 9.6	Predicted impact of new district pool on seven existing pools using overlapping catchments and attractiveness	295
Table 9.7	Predicted level of use at seven existing pools after location of a pool at Comiston as a percentage of the original value	298
Table 9.8	Indices of accessibility based on overlapping catchments and attractiveness for two new pools at selected sites given as percentages of the values with new pools at W. Hailes and Crewe Toll	301
Table 9.9	Predicted impact of two new district pools on seven existing pools using overlapping catchments and attractiveness	303

SECTION I

The economic and political background
of the problem of location

CHAPTER 1

The background and scope of the work

Introduction

It is conventional to say that in western society resources are basically allocated by two processes: private goods through the market place, and public goods ultimately through the political forum. Many of the wants and necessities of modern life in city and country are now met by the public sector. Writing of the United States in the late 1960's Teitz focused attention on the fundamental importance of public services:

"Modern urban man is born in a publicly financed hospital, receives his education in a publicly supported school and university, spends a good part of his life travelling on publicly built transportation facilities, communicates through the post office or the quasi-public telephone system, drinks his public water, disposes of his garbage through the public removal system, reads his public library books, picnics in his public parks, is protected by his public police, fire and health systems... Ideological conservatives notwithstanding, his everyday life is inextricably bound up with government decisions on these and numerous other local public services."

(Teitz, 1968, p.36)

However, as a synthesis of research edited by Berry and Horton in 1970 revealed, the overwhelming emphasis of work in urban geography till then had been on the organisation and

distribution of private services, especially shopping; public services had been relatively neglected.

An obvious yet striking feature of the list of services cited by Teitz is that many are only obtainable at a few locations in space. Several, for instance libraries and clinics, can be characterised geometrically as focal points where flows of users from surrounding catchment areas converge. For a few others, fire services for instance, this movement is reversed and the service is delivered from the central point to outlying areas. In both cases the service is supplied at a system of points whereas demand or need is continuously distributed over the area concerned. Such a service can be viewed as a system which is organised in space; its spatial form partly reflects the way the system attempts to fulfil its obligations to the public. This 'spatial organisation' is therefore likely to have a significant bearing on how well the system satisfies the needs of the population it is designed to serve.

The present thesis will be concerned with those services where it is reasonable to assume that the governing principle in planning the location of supply points should be to maximize the convenience of users. Where this assumption is appropriate, the supply points may be referred to as 'central facilities' and the problem of locating them is generally called the 'central facility', 'location/allocation' or 'multiple location' problem. Though they are relevant to many other services, the concepts discussed and developed here will be applied mainly to public swimming pools, which clearly fall into the category of central facilities. It may be appropriate to note at this point that, in a review of

recent work in economic and urban geography, King (1976) suggests that work on problems of locating facilities can help to initiate more effort on problems in the public sector.

Geographers and other location analysts (Christaller, 1966; Lösch, 1954) have developed a number of theoretical frameworks for understanding the spatial organisation of central points in relation to their surrounding market areas. Formulated with the operation of market forces in mind, these models have almost exclusively been applied to private commercial and retailing services and involve rather restrictive spatial assumptions. Nevertheless, some notions derived from central place theory, for example the concepts of demand cones and service thresholds, may be quite helpful initially in examining public services because they focus attention on the spatial relation of supply and demand. Indeed, Smolensky, Burton and Tideman (1970) have adopted Lösch's notion of a demand cone to describe the relation between access and effective demand for a central public service. Dear (1974), however, has argued that a conceptual framework for treating the problem realistically must take account of the distinctively public nature of the decision process involved. Even setting aside their restrictive assumptions about the distribution of demand, this is a requirement which central place models, as formulated at present, cannot meet.

The pioneering work of Teitz (1968) probably represents the first attempt to formulate a genuinely distinct theory of location for public facilities. While he presents a very neat set of constructs depicting the relation under a budget constraint between the scale and number of facilities, the operating costs of the

system and the quantity consumed, Teitz admits himself that he evades the purely locational aspect of the problem to a large extent. Furthermore, although his framework incorporates several distinctive features of public goods, it does not explicitly take account of the political element involved in decisions in the public sector. As a preliminary to considering more explicitly the spatial aspect of the problem, it is now convenient to examine how public services may differ in nature and locational organisation from private services.

Some general characteristics of private and public services

It is helpful to begin this discussion by sketching two highly simplified pictures of how goods and services are allocated in the two sectors. According to the neo-classical model of the private sector, provided income is not too unequally distributed, the aggregate priorities of individuals are expressed in effective demand: through the price mechanism, a perfect market then ensures that these preferences evoke an appropriate and prompt response in the right location. With due allowance for some difference in the time scale of responses and other serious differences such as the possible formation of parties, the notion of consumer sovereignty operating through the voter-politician relation instead of the consumer-entrepreneur relation can be made to yield a similar picture of the public sector. In the ideal model of a democracy with full participation and with political power equally distributed between individuals (for instance, as envisaged by Jefferson) the state and its elected representatives might be sufficiently sensitive to the public's priorities for the allocation of public goods to accord quite closely with the

aggregate wishes of individuals in respect of both kind and location. Using this approach, Tiebout (1956) argues that individuals will tend to migrate to areas where the local authority provides the combination of services and taxes they prefer. Tiebout's approach may have some relevance to American cities but it obviously has less relevance to the large unified authorities which now govern urban areas in Britain.

A similar notion of equilibrium regulated by the decisions of a host of genuinely independent actors underpins the model for the public as well as the private sector. Within the confines of resources available, according to both models, the most important material wants of most actors would be catered for, with the possible exceptions of some minorities.

The assumptions that excessive inequalities of income and power do not exist and that the actors are in some sense independent may be formidable ones. In these important respects urban industrial society may be much further from the idealised market model than the world of small units in farming, trade and manufacture which Adam Smith originally had in mind (Galbraith, 1974; Watson, 1976). Galbraith in fact argues that in the U.S.A. inequalities in both sectors interact very strongly:

"The unequal development is unrelated to need; the inequality in income bears no necessary relation to productivity or efficiency. Both are the result of unequal deployment of power."

(Galbraith, 1974, p.X)

In consequence, Galbraith argues, both equilibrium models serve more to conceal the underlying economic realities than to reveal them.

He goes on to claim that the market model may help to prevent the individual in the U.S.A. from seeing how he is governed and how resources are really allocated, calling it "the cloak over corporate power". Thus two very contrasting pictures of how economic and political realities are intertwined can be produced.

It is interesting to note briefly that some of Galbraith's arguments find a geographical counterpart in studies of the process of political decision-making involved in locating particular facilities. Thus Wolpert, Deer and Crawford (1975) remark that 'noxious' facilities such as sewage plants and incinerators tend to find their way to areas where the local community is least able to oppose the decision. By demonstrating how important the unequal deployment of power can be, such studies illustrate clearly the limitations of Tiebout's approach.

In fact, any decision to locate a public facility is effectively a decision to distribute certain benefits and costs among different groups of people (Austin, Smith and Wolpert, 1970). If the facility is seen as a desirable resource for people to use (e.g., a park or library), adjacent residents will enjoy improved convenience of travel and quite possibly more frequent use and improved amenity in their neighbourhood. With an undesirable or noxious facility, needed by the city as a whole but not of particular use to the local community (e.g., a sewage plant), people living nearby are likely to suffer loss of amenity and a drop in the value of their homes may well result. It is worth noting that in both cases the strength of the benefits and costs involved will generally be quite closely related to the degree of proximity and are therefore amenable to geographical treatment.

Much of the public conflict around planning decisions involves attempts by particular groups to influence the incidence of such imponderable negative and positive benefits (Harvey, 1973), especially to prevent undesirable projects or shift them elsewhere. The final decisions about such projects are essentially political decisions. As noted earlier this feature of public services is one important difference vis-à-vis services in the private sector and some of its consequences will be explored later.

A striking peculiarity of certain goods supplied by the public sector is that, once supplied to one individual, they are thereby supplied free to all. Broadcasting, lighthouses and city streets, are examples of this. Such goods are inherently unsuited to rationing through the normal price mechanism mainly because users cannot be charged according to the amount consumed. In welfare economics these are called 'pure public goods' (Webber, 1973). Though many central facilities are free at the point of supply itself, all of them involve transport costs for users and can therefore be classified as 'impure' or 'semi-public goods'.

Fire stations and most public libraries fall into the latter category. As most public swimming pools in Britain have entry charges, they cannot be classified as 'semi-public' goods. Nevertheless, a decision about the location of a pool is a political decision about how to distribute certain benefits and costs. Furthermore, since most public swimming pools in Britain are strongly subsidised (Scottish Sports Council, 1979) and entry charges are relatively low (25 pence for an adult, 12 pence for a child at most pools in Edinburgh at the time of writing), for many users the cost of travel is a very significant part of the cost of

using the facility. On the other hand, for retailing and other private services the cost of travel is usually only a small fraction of the total cost of obtaining the goods or service concerned. In this respect swimming pools have more in common with libraries and fire stations than with private services.

The spatial organisation of public services

Some general features of public services have been noted briefly. The question arises whether the distinctive characteristics of the public sector have any specific impact on the location of central facilities. Various aspects of this general question can now be explored.

(a) Lack of Competition

Since competition from alternative suppliers does not exist, an obvious first question is what result does this have. In central place theory, which is based on the principles of neo-classical economics, such competition ensures that retailers are located so as to maximise overall accessibility to consumers. Clearly competition should also provide retailers with a strong incentive to adjust their locations to changes in the distribution and mobility of population. Because the penalties of inefficiency are apparently weaker in the public sector and the criterion of profit and loss does not apply, it might be expected that, whereas a badly located grocer will go out of business, a badly located swimming pool will either remain underused or inflict excessive travel on its patrons. Conversely, since the incentives to enter the market are also weaker, a swimming pool for which there is sufficient local demand may not be constructed.

Experience does not always confirm these expectations as far as the private sector is concerned. Clarke(1975) notes that private services can lag as much as public services in moving to New Towns and new housing areas. Furthermore, Jones (1967) suggests that many retailing activities in Edinburgh have shown much caution in leaving the central area and that the decentralisation of population since 1945 has not, apparently, evoked an equivalent response from shopkeepers. Of course, there may be reasons other than inertia for this failure to leave the centre. The fact that bus routes still focus on the city centre and the advantages to shoppers of comparison buying and of being able to make multiple purpose journeys may help to keep the city centre attractive to retailers. In the absence of detailed research it is therefore difficult to say how important inertia is in this case.

In contrast, some services in the public sector seem to adapt fairly quickly to changes in population and it can be shown that some public services in Edinburgh are rather efficiently organised in space. If the distance travelled when each member of the population makes a trip to his nearest library in the city is measured, the total distance travelled to the eighteen libraries in existence in 1975 is only 13.6% greater than it would be if the libraries were located to minimize this travel distance. If a similar measurement is made from fire stations to areas containing population, the equivalent difference is only 12.9%. By its nature the fire service is bound to be specially conscious of accessibility. Indeed, a more appropriate measure of spatial efficiency might well reveal the fire service to be even better located than this figure suggests.

Even when the special circumstances of the fire service are borne in mind, these two examples suggest that centrally co-ordinated locational decisions are capable of a satisfactory level of spatial efficiency, whether this is achieved or not. Moreover, both of these services would seem to have responded more promptly to changes in population than the retail services examined by Jones. Part of the explanation may be that public services are less tied to the centre by the advantages of comparison buying. It is difficult to invoke bus routes as part of the explanation because it is hard to see why these should influence private services but not public ones. Thus the absence of competition is not so detrimental to spatial efficiency as might be expected and may in fact facilitate greater efficiency in some circumstances.

The reasons for thinking that public services are able to locate more efficiently can be stated in a more general form. This argument derives from the distinction between 'competitive systems' and 'welfare systems'. In the competitive world of private services the rival suppliers will optimise separately: each will try to establish himself in a location which either brings him as much individual profit as possible (Scott, 1975) or ensures a satisfactory income with least risk or effort (depending on whether he is an optimiser or satisficer). As examples of welfare systems, public facilities, by contrast, are theoretically able to optimise the supply system as a whole because the locations of the individual centres can be co-ordinated to provide the most convenient configuration for the community at large.

Hotelling's (1929) well-known equilibrium solution for two suppliers on a uniform linear market affords one example of how, in

theory, competitive and welfare systems may differ spatially. Hotelling assumes that demand for the service is inelastic which means that the amount purchased by a consumer remains the same irrespective of his travel costs, i.e. it is independent of the location of supply. Under these conditions he is able to demonstrate that both rivals finally occupy the same central location to avoid being at a competitive disadvantage - a more costly outcome for consumers than when the vendors co-operate to space themselves evenly along the line. Hotelling extends this argument to three or more competing sellers, suggesting they would crowd together for similar reasons. By analogy he argues that competition leads to excessive similarity in kind of goods supplied as well as location.

It should be noted that with three sellers Hotelling's conclusion is debatable. Trial and error experiments suggest that with three or more sellers a stable pattern may not exist. In fact it could be argued that there may be some inconsistency in Hotelling's basic model: it assumes that consumers' demand is inelastic, i.e., in one sense insensitive to distance, but that all buyers patronise the marginally nearer vendor, i.e., in another sense they are very sensitive to distance. It would be unwise to draw definite conclusions from Hotelling's argument since it is based on a rather idiosyncratic problem. However, it serves a useful purpose in drawing attention to the difference between competitive and welfare solutions.

Hotelling's argument is unlikely to apply to a very large number of sellers on a uniform unbounded plane: there, even spacing would seem to be a stable arrangement since any supplier

who moves to encroach on the market of his neighbour will gain less than he loses to his neighbours in the direction opposite his movement. It is not clear whether Hotelling's argument applies to three or more sellers in the bounded space within a city. Hotelling suggested that the outermost sellers on a plane would tend to gravitate more towards the centre than public welfare would require, but the argument just outlined can be used to suggest that fairly even spacing may still be a stable arrangement. It is therefore hard to say whether Hotelling's argument can help to explain any observed differences in spatial efficiency between services in the public and private sectors or the apparently excessive concentration of retailers near the centre of Edinburgh. What is clear is that under these conditions the welfare system, in theory, will tend to match or surpass the competitive solution in overall convenience to users. It is interesting to note Scott's (1971) argument that, in two dimensions, competitive, welfare and monopoly systems may tend to have similar spatial forms. This suggests similar levels of efficiency for the three systems but Scott does not develop the argument in depth or cite examples.

It must now be strongly emphasised that the assumption of inelastic demand is crucial to the outcome in all of Hotelling's examples. If demand is assumed to be fairly elastic, the more distant consumers make less use of the service. Each point of supply then tends to draw most of its customers from a limited area nearby. In Hotelling's original problem clustering is then an unstable arrangement for the two vendors in competition because it means sharing a restricted local market; each seller will enjoy more custom if they are evenly dispersed along the line.

Dispersion is the welfare optimum as well as the competitive one, since it reduces the overall cost of travel and increases the overall level of demand. Thus when demand is elastic competitive and welfare optima on a line coincide.

A similar conclusion can be reached for a plane because more demand will be stimulated by a dispersed pattern of location, irrespective of whether the sellers are in competition or not. Of course, this pattern of dispersion under elastic demand forms the basis of Christaller's models. In fact it can be argued that the more elastic the demand, the stronger the tendency to disperse. Conversely, the less elastic the demand the less retailers tend to lose as a group by clustering. It is not clear what the competitive optimum with inelastic demand and many sellers on a plane will be but it can be argued that if clustering occurs it may persist for the reason just outlined, thereby rendering the competitive outcome less efficient spatially than the co-operative one.

Thus on a theoretical level it is not possible to draw any simple conclusions about the effect lack of competition has on location. In theory, a centrally co-ordinated service can at least match the competitive solution in its efficiency, regardless of whether demand is elastic or not. When demand is elastic, both types of system will in fact tend to assume the same dispersed form. On the other hand it is possible that some public services may be less responsive to dynamic factors such as changes in the distribution of population, but the library and fire services in Edinburgh do not seem to support this argument.

Finally it should be noted the preceding argument relies on

the Euclidean properties of a continuous space. In a strict sense the location of all services in the city should be studied in relation to the complex network formed by the city's transport system, particularly bus routes. In practice, however, this would be difficult because of the large amount of data required and the time needed to compile it. As far as a theoretical examination is concerned the difficulty is that theory of the kind developed by Hotelling and Christaller for continuous spaces has not yet been developed for networks. It is therefore very difficult to say whether consideration of networks would reveal a greater tendency to cluster or a greater difference between welfare and competitive optima. Since networks can vary enormously in structure, generalisations of this kind may well be impossible and the tendency to disperse or cluster may depend on the particular network.

(b) Accessibility as an external benefit

At the point of use, many public services are free or have relatively low entry charges. As a result travel cost, including time and inconvenience, often represents the major real cost to the consumer (Teitz, 1968; Smolensky, Burton and Tideman, 1970); the price paid by users or a significant part of it is therefore largely a result of the form of spatial organisation adopted by the supplier. Consequently, an increase in the number of facilities may increase overall demand by reducing price.

Within a budget constraint, however, more facilities will mean smaller ones and in some instances (e.g., libraries) smaller facilities may be less attractive. As Teitz notes, in terms of attracting demand a trade-off then exists between scale and number,

so that public services are in the curious position of being able to generate or inhibit demand by organising themselves appropriately since demand and supply are tightly interlocked through the spatial form of the system. It is interesting to note that this could be used as a partial explanation for the phenomenon of 'latent demand', noted in recreation studies (Coppock and Duffield, 1975).

If the size and number of facilities is given, and the facilities are regarded as desirable, it seems reasonable to assume that access to the service is the major benefit being distributed in any decision about their location. Under these conditions maximizing overall accessibility, however defined, will maximize the total social benefit from a particular project.

However, maximizing total benefit on its own ignores the social distribution of costs arising from the differing incidence of local and national taxes raised to finance the service. With this point in mind Dear (1974) has formulated an elaborate model capable of dealing with the differential social and spatial incidence of the respective costs and benefits of financing and locating facilities, but admits that it is very far from being operational. Dear's model may be conceptually useful, especially in the case of noxious facilities, but the latter are excluded from the present study. One cannot deny that the incidence of costs for desirable facilities varies in a similar way but they are ignored here on the principle that central facilities should be sited only with respect to potential users. After all, taxpayers are hardly likely to use a public service in proportion to their tax contributions. Other costs to residents arising in the immediate vicinity of a facility (e.g., through exacerbated parking problems near a swimming

pool) are also ignored, because the emphasis here is on the question of situation, not siting. Besides, such problems can often be alleviated by appropriate design of the facility concerned (Dear, 1975).

Because the benefits conferred by improved access are not directly priced, they can be regarded as a kind of windfall gain or external benefit. Such externalities, as they are sometimes called, are a very pervasive feature of urban life (Harvey, 1973; Mishan, 1967). It can be argued that they will eventually be priced through location rent in the housing market, but the siting of a new facility will often come to the sitting occupant as a windfall gain or loss to be realised only when the house is next sold. Moreover, a study by Dear (1975) in Philadelphia found no change in property values after the siting of a mental health facility which the local community had opposed.

Externalities are quite familiar to geographers in a somewhat different guise. In drawing the traditional isochrones of travel time or distance-decay fields around recreation centres geographers have in fact been describing the strength and spatial incidence of externalities, though they may not have viewed it in this way. As noted earlier, this indicates that some externalities are intrinsically geographical in nature.

The fact that negative externalities may not be fully compensated partly explains why proposed roads provoke such strong opposition. In contrast planning proposals involving positive externalities rarely generate so much controversy. Rival neighbourhoods seldom contest the siting of a library quite so conspicuously. On the whole, disputes about the location of external benefits are

usually mediated through planning machinery rather than the market, an acknowledgement of the unsuitability of the market for dealing with such problems.

It can be argued, however, that the political forum has no reliable means of obtaining a generally favourable resolution of such conflicts, whereas, in theory, the market has. For instance, facilities may be located to suit politically powerful groups to the detriment of the community in general. It may therefore be useful to devise formal models which take account of the incidence of costs and benefits to the society as a whole and try to use these models to find socially optimal solutions. Such models cannot, of course, be a dependable means of yielding an ideal answer to the problem. They may, however, facilitate an understanding of the problem and an assessment of some of its components, thereby providing useful information to the political/planning process. The examination, development and application of such models is the central concern of the present thesis.

The absence or relative unimportance of entry charges also means that the normal market criteria for investment in extending the capacity at a particular site are absent, though statistics on users may be a useful substitute. The related problem of deciding between the extension of an existing facility and the construction of a new one some distance away remains impervious to both criteria, and is common to private and public sectors. It is interesting to note that by providing an assessment of how much demand a new facility will attract location models may help to resolve this type of question in the private as well as the public sector.

(c) Influence of the political process

There are other features of public services which derive more directly from the political nature of the decision process. Some of these have been briefly mentioned already but it may be helpful to elaborate them further. Certain facilities (e.g., discothèques) may have costs to residents in the immediate vicinity but bring benefits to patrons who come from a much wider area. This means that they have both positive and negative externality fields with very different spatial extents. The users may thus be a different group from the locally affected residents and a conflict of interest may arise between the two groups over the location of the facility, even if the local residents are also users. Wolpert, Mumphrey and Seley (1972) provide some examples of services which respondents wished to have within convenient reach but not immediately adjacent. Bus stations provide a good example.

Such cases present a difficult problem for planners and local politicians. The problem may arise when the facility is a relatively specialised one needed only by a small minority of residents. Dear's (1975) study of drug treatment centres in Philadelphia and the work by Wolpert, Dear and Crawford (1975) on 'community mental health satellites' in California provide acute examples of this dilemma. Both studies suggested that the opposition of the better organised neighbourhood groups was so effective that planners in health care tended to adopt a strategy of locating the facility where it was easiest to overcome local opposition. As a result, facilities were frequently placed in areas of the inner city which were often far from being the most therapeutic environment.

The present study excludes noxious and controversial facilities

and therefore avoids those facilities where the political process is normally most active. Also the services considered are not specialised but can be assumed to be of general benefit to a substantial section of the population. It seems reasonable therefore to assume that positive external benefits are a function of access. Special interest groups may still be able to influence the location of such facilities and this has to be borne in mind when actual decisions are examined. In addition, the outlook of the various professional groups involved in planning and administering public services may sometimes have a bearing on the final decision.

'Each profession is a conspiracy against the laity.' In a discussion of the way public goods and services are planned, Webber (1973) points out that this dictum of Shaw's anticipated the findings of a number of recent social researchers. With the U.S.A. mainly, but not exclusively, in mind Webber goes on to argue:

"The professional shares the behavioural and value norms of middle-class culture, and middle-class culture is simply different from the non-middle-class cultures of other groups. The working-class ... , who have inadequate social skills for dealing with the majority culture, thus have difficulty in breaking through the cultural barriers that surround hospitals, schools, housing administrations, and the like. When income deficiencies further reduce their capacities ... their handicap ... is compounded."

(Webber, 1973: p.58)

An instance of this line of argument being carried much further is

cited by Wolpert, Deer and Crawford (1975) who state that some local groups regarded community facilities for mental health as a "middle-class rip-off", meaning that they were of more value to the providers of care than to the recipients. Certainly the administrators of a service are not an entirely disinterested group (Levy, Meltsner and Wildavsky, 1974).

Though politicians, on behalf of the electorate, are ultimately responsible for public agencies, it is hard to believe that the professional officials who staff these agencies have no influence on the nature and location of the services supplied. Should professionals unconsciously favour middle-class clients, the absence of a market and the presence of cultural barriers may help to conceal the needs of other groups. Whatever its faults, the market in theory does make suppliers accountable to the consumer whereas planners are rarely so directly accountable to a broad spectrum of the planned. On the other hand, professionals may sometimes be able to redress the bias which the planning process has against the less powerful and less articulate through disinterested assessment of need.

Thus, the nature of the political process and the various agents involved in it may influence decisions about the location of central facilities, favouring certain areas or social groups and prejudicing others. This distortion provides part of the rationale for developing normative models since these can help to identify locations which are optimal in terms of the population of the area viewed as a whole.

Before examining existing models of locating central facilities and attempting to develop some new models it is convenient to

outline the rationale for developing such models:

- (a) They provide a yardstick against which proposed locations for new facilities can be assessed. Similarly, by helping to detect bias in the existing or the proposed locations, they provide a possible antidote to the imperfections of decision-making in the public sector.
- (b) The existing location/allocation models were largely developed to solve location problems such as the location of plants and depots, mainly in the private sector. Such models invariably assume that demand does not depend on the location of facilities, whereas Teitz has argued that this is often not true in the public sector. Clearly there is a need to develop models to suit those services in the public sector where demand is spatially elastic.
- (c) Since this field of research has only developed fairly recently, existing models are often rather simplistic. For instance, the catchment areas of facilities are usually defined by Thiessen polygons, an unrealistic arrangement since it does not allow catchments to overlap. Hence there is a need to develop more realistic models.
- (d) On a more theoretical level, neo-classical theories of location have mainly been concerned with competitive situations on a uniform plane. By treating non-uniform planes and networks and by identifying welfare solutions, such models may help to broaden the range of location theory.

Formal structure of the problem

The preceding summary provides a convenient introduction to the formal definition of the problem. In all models, distance (as a surrogate for access) is presumed to reduce the value or usefulness of a facility to the consumer and can therefore be called a disutility. In developing normative models the following data will be assumed:

- (1) the location of each demand point on a plane or in a network;
- (2) the requirement or demand at each of these points;
- (3) the capacity of facilities, where appropriate;
- (4) a suitable way of measuring the disutility of distance.

In optimizing the spatial organisation of facilities the basic elements which can be varied are:

- (1) the number of facilities;
- (2) the location of each facility;
- (3) the allocation of demand points to facilities (and thereby the size of the facilities).

These data and variables define the general location/allocation problem.

It is useful to define a notation for representing the problem mathematically:

- n will denote the number of discrete demand points;
- m will denote the number of facilities;
- i will denote a particular demand point;
- j will denote a particular facility;
- d_{ij} is the shortest distance from i to j measured according to a plane or network metric;

p_i is the demand at i expressed as a number of trips per time interval, normally taken here to be directly proportional to total population.

In problems of locating depots, the depots are usually called 'sources' and the demand points 'destinations'. Here movement is usually in the opposite direction so these terms would be confusing, although the structure of the problem is not affected.

The role assigned to capacity is likely to depend on the individual service. Toregas and ReVelle (1972) have argued that many public services effectively operate without real constraints on capacity. Of course, all facilities do have a limit to their capacity, but Toregas and ReVelle apparently believe that they rarely operate at the point where capacity is so fully utilised that arriving customers have to queue or go elsewhere. This argument certainly holds for libraries and for swimming pools much of the time.

When constraints on capacity are used in a model, customers are normally diverted to the next nearest centre once a particular facility is full. Gould and Leinbach (1966) used such a model for locating hospitals and adjusted the capacities of the hospitals through a series of iterations which progressively removed unnecessarily long journeys by patients. Yeates (1963) also used constraints on capacity as a means of delineating school catchment areas which would minimize the cost of travel in the system, given the existing sizes of schools in the study area. Since schools and hospitals normally operate near to the limits of their capacity, it seems appropriate that the size of such facilities should influence both their location and the extent

of their service areas, as in these two examples. In fact, constraints on capacity are probably more common than Toregas and ReVelle suggest but no one has tried to outline which services fall into this category or provide a systemic basis for defining the role of capacity. Work by Öberg (1976) and McCalden (1981) suggests that capacity has an important influence on the access of patients to a dentist.

It is interesting to recall Teitz's suggestion that in services where consumers have a choice, the size of a centre may be a measure of its attraction. Using this as a point of departure, ReVelle and Church (1976) develop a model which makes Teitz's formulation of the problem mathematically explicit and also outline a procedure for maximizing utilization of the service under a budgetary constraint. To make this model operational two critical parameters - exponents for distance decay and size attractiveness - are required. The latter parameter may present quite a difficult problem of measurement.

If no constraints are placed on the capacity of a facility and if size is not used as a measure of attraction, this is equivalent to assuming that the number of people served by a centre and the areal extent of its catchment depend entirely on its location. Conversely, when capacity is specified in advance, location has no influence on the size of the facility; instead size influences location. If users are free to choose facilities and if constraints on capacity generally do not affect the user's access to the service, it seems preferable to develop models where size depends on location. Since both of the latter conditions seem to apply to such services as libraries and swimming pools, constraints

on capacity are omitted from the models discussed here. Nevertheless, it can be argued that with both services size of facility should be used as a measure of attraction, a point which will be discussed in more detail later.

The approach used therefore tends to yield small facilities in areas of low density and large facilities in areas of high density of population. Although this procedure carries the danger of occasionally allocating facilities which are too small to be practicable, centres could be required to be larger than a specified minimal size by adding an appropriate constraint to the model used. In the present study of swimming pools in Edinburgh, however, this difficulty never arose.

Thus the basic elements considered in the models presented here are the number of facilities and their location. The problem of optimizing the number of facilities will be discussed in section two; section three will be concerned with finding optimal locations.

The Role of Access

Having outlined the role of optimizing models and defined their basic elements, the goals to be optimized now have to be specified more precisely. Accessibility to population is the main goal to be optimized, though it would perhaps be more accurate to say that the goal is to minimize the effects of inaccessibility. For the sake of simplicity in the following discussion access is equated with distance.

In defining the effect of distance the distinction, noted earlier, between models based on elastic and inelastic demand is crucial; yet it has received insufficient attention in the

literature. In models based on elastic demand the individual's demand or use, expressed in trips to a centre, depends on the price of obtaining the service. Where public services allow users to enter without a charge (e.g., public libraries, museums, parks), the price of the service can be taken simply as the travel cost. Where there is a standard entry charge (e.g., for many swimming pools in a city) the spatial variation in price is essentially a function of the variable travel cost, the other component of price being constant. In both cases the demand from population unit i will then depend on its level of access and therefore on the location of service centres (Smolensky et al., 1970).

It is important not to confuse elastic demand with the well-known distance decay effect around individual facilities. A simple example may help to clarify this distinction. Consider a certain service with inelastic demand which is met by four facilities in a city. The number of trips from particular population subareas may then vary with such factors as age, income and social composition but will be invariant with respect to the location of the centres. Suppose a certain proportion of users from any subarea, i , go to their nearest centre, and successively smaller proportions go to the second, third and fourth nearest facilities. Around any facility, j , the density of trips (i.e., the proportion of users who prefer to use j) will be higher near the facility and much lower for areas further away where, for instance, j may only be the fourth nearest centre. A study of the catchment area of one of the facilities will then reveal a distance decay effect, though demand is inelastic spatially. In short, if demand is inelastic, distance decay describes how

distance affects choice of centre; whereas with elastic demand models, distance decay firstly describes how demand falls with distance. In the latter case a second decay effect may subsequently be used to describe the choice of centres.

Although several studies report distance decay effects around supply points (Berry, 1967; Taylor, 1971), their design often does not permit an assessment of the extent to which this represents a fall in demand as opposed to a decline in preference. This question can only be approached by comparing the patterns of use of individuals with high and low access through a household survey or through a survey of users at all facilities in an area. The results of a study of the latter kind by Weiss and Greenlick (1970) suggest that, for medical services, distance mainly affects the choice of facility rather than frequency of use. A study by Gainey (1977) broadly supports this conclusion.

On the other hand the cartographic evidence of an exploratory study of the use of public swimming pools in Edinburgh (Currie, 1977) tentatively supports the argument that demand for this service is at least moderately elastic. A major study carried out by the Scottish Sports Council (1979) of the use of various swimming pools in Scotland maps several catchment areas in detail and throws much light on various factors influencing their extent, but fails to investigate whether demand is spatially elastic. Part of this study involved a household survey to find out why many people do not use swimming pools but unfortunately the opportunity to use this survey to compare access and frequency of use in a suitable manner was not taken.

The nature of demand for some services, however, seems

relatively clear cut. For obvious reasons it seems sensible to assume that demand for fire services, for secondary education and for the most essential medical services is strongly inelastic. In such cases it would be just as appropriate to speak of 'need' as of demand. The concept of 'supply-led demand' (Coppock and Duffield, 1975) often quoted in relation to recreation services is equivalent to postulating elastic demand. Clearly more research is needed to determine how elastic is the demand for those services which are not so easy to classify. The conclusions of such studies will have a significant bearing on what kind of objective function is appropriate for a particular service. Of course in such studies there is always a danger of confusing demand, which is partly a function of income, with need, which is not dependent on income but harder to define.

When demand is elastic the obvious goal is to locate centres to maximize the level of total use or demand generated. If demand is a simple inverse function of distance and only one centre is being located this goal might be expressed:

$$\text{Maximize } z = \sum_{i=1}^n p_i / d_{ij}^b \quad (j = 1)$$

where the exponent, b , measures the decrease in use with distance.

Where demand is inelastic, the cost of obtaining the service still increases with distance. One appropriate objective is then to minimize the inconvenience of travel in the system by minimizing aggregate travel distance as follows:

$$\text{Minimize } z = \sum_{i=1}^n p_i d_{ij} \quad (j = 1)$$

Both of the preceding objectives are based on efficiency. It can also be argued that a location should be chosen to distribute the cost of travel equitably among the population served. Some inequality in access is inevitable because some people will always be nearer the point of supply than others. One way of finding a solution based on equity is to select a location which reduces the journey of the most distant consumer to a minimum, often called the 'minimax' solution. This principle accords with Rawls's (1972) criterion of justice whereby the 'prospects of the least fortunate are as great as they can be'. Formally this principle can be expressed as:

$$\text{Minimize } z = \text{Max } |d_{ij}| \quad .$$

For certain services, particularly emergency fire and medical services, the value of the service to the user declines with distance from the point of supply. A desirable standard of service may then be defined in terms of a maximum time or distance, s , and the service agency may wish to position facilities to ensure that the whole population is within s units of a centre. A point within s units of a facility is then said to be covered. For any potential site j a binary coefficient a_{ij} can be used to describe which demand points can be covered from j ; when i is covered, a_{ij} has the value one; otherwise it is zero. If one facility is to be located so that the number of people it covers is maximized, the objective can then be written:

$$\begin{aligned} \text{Maximize } z &= \sum_{i=1}^n a_{ij} p_i \\ \text{where } a_{ij} &= \begin{cases} 1 & \text{if } d_{ij} \leq s \\ 0 & \text{if } d_{ij} > s \end{cases} \quad . \end{aligned}$$

If the value of s is relatively large this objective may approximate to the minimax solution. Indeed, it could be used as a more flexible means of incorporating an element of equity into the solution of the location problem.

In broad terms it can be said that minimizing overall distance will tend to increase use, reduce cost, improve equality of access and, where relevant, increase the number of people within a certain radius of cover. More rigorously, however, each possible goal has a different mathematical form; the location which maximizes use will therefore rarely co-incide exactly with the point of minimum cost. Hence each goal will produce a somewhat different solution. The selection of a goal to use in siting a particular facility is partly an empirical question about what assumptions are appropriate but, since this choice also implies striking a certain balance between equity and efficiency, it is also partly a social value judgement to be made politically.

When the existing literature on location/allocation models is examined with this range of objectives in mind it is clear that inelastic models have dominated the field. Thus, although Massam (1975) provides the most comprehensive geographical text on the subject to date, apart from a brief mention of a covering model he only discusses one objective, that of minimizing aggregate travel. Moreover, he fails to mention this model's assumption that consumers have a fixed requirement for the service which is completely independent of the location of supply. A discussion of 'distance decay' around centres is included in an earlier chapter, but the difference between elastic and inelastic demand is not mentioned. In a more recent work Massam (1980) mentions maximum utilization as a possible goal but does not discuss any model

with this objective.

Massam's treatment reflects the bias of preceding work which almost invariably concentrates on this one objective without mentioning the attendant assumption. An invaluable monograph, which documents several computer programs for solving location/allocation problems (Rushton, Goodchild and Ostresh, 1973), shows a similar concentration on cost-minimizing models; perusal of a very extensive bibliography on the subject by Lea (1973) confirms this impression. A similar point can be made about a very useful review of methods of optimization by Scott (1971) and about the monograph by Törnqvist, Nordbeck, Rystedt and Gould (1971). Thus, Teitz's (1968) concept of locating centres to maximize use has largely been neglected as far as operational models are concerned.

In problems of locating depots the requirements of each destination are usually taken as fixed, so a situation involving elastic demand is, presumably, unlikely to arise. A review of such models by Eilon, Watson-Gandy and Christofides (1971) illustrates this general point quite well. The heavy emphasis in the geographical literature on minimizing the cost of travel may therefore be a carry over from such work, mainly found in periodicals covering the field of operations research. Another reason might be that problems with elastic demand are harder to solve or at least appear so. Nevertheless, given the preoccupation of geographers with distance decay as a general phenomenon, it is surprising that the idea of use falling with distance has not been incorporated sooner in models for locating facilities.

An interesting exception in operations research is provided by Abernathy and Hershey (1971), who formulate a model based on elastic

demand in a paper concerned with planning the location of regional centres for health care. The authors discuss results from a hypothetical example but it is not clear whether the program employed is able to search an area comprehensively or just evaluates certain specified points. The notion of elastic demand is also incorporated in models for maximizing the joint welfare of consumers and producers formulated by Wagner and Falkson (1975), but their models are as yet purely theoretical and the authors do not attempt to make them operational.

Although models based on elastic demand have been neglected, the objective of maximizing the population covered by the system of centres has received a fair amount of attention. Toregas and ReVelle (1972, 1973) and White and Case (1974) develop a series of such models and provide a number of applications. Attempts have also been made to develop models which maximize equity, for instance Morrill (1974), Morrill and Symons (1977), McGrew and Monroe (1975) and McAllister (1976).

Optimizing versus satisficing models

It might be argued that satisficing models based on the concept of 'bounded rationality' propounded by Simon (1957) would be preferable to optimizing models because they correspond better to the way in which decision makers in both the public and private sectors actually behave. Undoubtedly satisficing models would give a more realistic description of locational decisions and the levels of efficiency likely to be attained in practice. The aim of the present work, however, is not to replicate reality but to try to develop methods of improving reality by finding better locations than might result from a typical process of decision making.

Of course, it is unrealistic to expect locational decision makers to find and implement the "one best" solution. Several imponderable criteria will often have to be weighed against those which can be measured satisfactorily enough to be included in the optimizing model. Nevertheless, the two main variables in any decision about the location of a facility are likely to be or ought to be demand and access. These are both measurable and form the basis of the optimising models.

Moreover, as Eilon (1972) has demonstrated, the two approaches are not so radically different as first appears. The satisficer wants a solution which is "good enough" on certain criteria; for instance, it might be below a certain cost or above a certain level of revenue. The satisficer's targets are therefore in the form of constraints; these constraints define a set of feasible (i.e., satisfactory) solutions. Similarly, the constraints in optimizing models define a set of feasible solutions. If the constraints are linear both feasible sets form a convex group of points. The difference is that, with the latter, one overall objective, such as minimizing costs, can be specified to select the best solution from within a feasible set. In short, a satisficing model can be viewed as equivalent to an optimising model with no objective function; in Eilon's words,

' ... in satisficing there is no difference between goals and constraints.'

(Eilon, 1972, p.7)

Eilon's paper draws attention to the flexibility of optimal programming models in allowing goals to be expressed in two forms: in the objective function, or as constraints. This sometimes

permits the interchange of constraints and objectives to form a model with a different emphasis. Thus, where a system has several goals which cannot be converted into compatible units to form one overall objective function, some goals may be expressed as constraints and optimized according to one particular objective function. This, incidentally, may offer a possible long-term hope for developing models which allow the goals of efficiency and equity to be pursued in tandem.

Some confusion in terminology may result because the term goal has generally been associated hitherto with the objective function. Here 'goal' will refer to any formal aim whether expressed as a constraint or as an objective function. 'Objective' will refer only to the latter.

The construction of satisficing models would probably entail a study of the decision-making process and the decision makers themselves; the problems and practical difficulties of such an approach are sufficient to require a separate study. The main reason for concentrating on optimising models, however, is that by providing a basis for evaluating actual alternatives and proposals, they are likely to be of more direct value to policy makers.

Focus and organisation of the thesis

The main concern of the present work is with concepts, models and methods relevant to solving the problem of organising systems of central facilities. There is a particular emphasis on developing new methods and models. Some of the models developed will be applied to the problem of locating swimming pools in Edinburgh partly as a means of assessing their value in locational analysis,

but analysis of particular services is not the primary aim and so forms the smaller part of this work.

However, a series of reports has already been published on various recreational facilities in Lothian Region including squash, swimming, bowls and golf in which the emphasis is on applying some existing models to help devise a strategy for future provision (Cargill and Hodgart, 1977; 1978).

The main justification for developing models, of course, is their potential application. The ultimate practical aim is therefore reflected in a strong bias towards models which have a reasonable chance of being made operational. Models which yield only conceptual insights receive less attention.

The focus on models and concepts grew out of empirical work on services in Edinburgh in which the cost-minimizing algorithm developed by Törnqvist (Törnqvist, Nordbeck, Rystedt and Gould, 1971) was used exclusively. During this work it became clear that this method was based on assumptions which limited its application more severely than realised initially and it therefore became desirable to develop new approaches which would be less restrictive. In doing so, particular attention was given to models involving elastic demand partly because of previous neglect.

Having defined the problems to be examined and the main emphasis of the work, it is now possible to outline how the rest of the thesis is organised. The main content can be divided into three parts. The first is concerned with developing a theoretical framework which allows questions relating to the optimal number of facilities to be examined. To this end, the relation between the cost of supply and the number of facilities provided is examined in

Chapter 2 and an attempt is made in Chapter 3 to estimate how users benefit when different numbers of facilities are provided under conditions of elastic and inelastic demand. By comparing these costs and benefits in various circumstances, it is possible to draw some general conclusions about the optimal number of centres for a given area in Chapter 4.

The next section is concerned with choosing optimal locations for a specified number of centres. Some existing models for locating centres in a network are discussed in Chapter 5. Chapter 6 is concerned with examining existing methods of searching space on a plane and with developing new methods for use with two objectives which assume elastic demand. A computer algorithm which implements the latter methods is then described and tested in Chapter 7. A particular feature of this algorithm is its ability to employ overlapping catchment areas.

The final main section, comprising Chapters 8 and 9, is concerned with the location of swimming pools in Edinburgh. In order to illustrate some of the factors which influence such decisions in the real world, the history of decisions about the location of two new pools is first discussed. The various optimizing models discussed in Chapter 6 are then applied to assess how far each decision departs from the theoretical optimum. An attempt is also made to assess the value of the models used for planning the location of facilities like swimming pools. The final chapter outlines the main conclusions of the work and suggests that certain aspects of the problem need further research.

The Time Dimension and Other Assumptions

Several simplifying assumptions underlie the models to be discussed. As far as the time dimension is concerned, no allowance is made for temporal variations in demand or accessibility. Indeed, it would be very hard to do so without constructing a simulation or dynamic programming model - probably a lengthy task. Where short-term fluctuations in demand may influence decisions about the capacity required, this factor may be handled by means of queuing theory (Massam 1972 and 1975). ReVelle, Marks and Liebman (1970) highlight the nature of this problem:

"Demands ... exhibit seasonal variation, day-night variation and variation due to economic conditions and weather conditions - factors which are difficult to predict. To design for the most severe case may be to misallocate valuable resources, but to design for the average may produce a system that fails under a stress situation ... "

(ReVelle et al., p.693)

This essentially stochastic factor is probably more relevant to the fire service than to any of the other services mentioned here. Thus, if many fires broke out in a very short space of time at scattered locations, a breakdown in fire service could occur. A similar problem could occur with emergency medical services.

ReVelle and co-authors also note the difficulty involved in planning the locations of facilities over time in the face of secular changes in population and technology. This aspect of the question can only be approached indirectly by the methods discussed here. The problem of locating facilities under various time

horizons can be examined through dynamic programming (Scott, 1971 and 1975) but one obvious difficulty with this approach is that the distribution of population may not be easy to predict.

The pattern of access in a large city can vary markedly during the working day. When most customers walk to a centre this variation may be ignored, but for certain services it may need consideration. If sufficiently important, it may be treated by simulation or by progressively altering the time/distance estimates used in the network form of the location problem.

Two further limitations stem from the partial nature of the analysis. First, the allocation of funds between different public services is not considered. This is equivalent to assuming a specified budget constraint on each service, a constraint which is common in the public sector. In this situation it is always difficult to know whether the budget is more or less than what would be optimal. Second, the locations for each service are optimised in isolation from other public and private services. Although this ignores trips by consumers for multiple purposes, it is not a completely intractable problem, especially in network models where feasible sites may be restricted to those nodes which facilitate such trips.

SECTION II

A framework for optimizing the number
of facilities

CHAPTER 2

Size of facility and a spatial supply curve

Introduction

In the previous chapter the nature and background of the 'central facility problem' were described; the basic elements of the system to be analysed were outlined; and a notation for representing the latter was established. This section, comprising Chapters 2, 3 and 4, will concentrate on one of those elements, viz., the number of facilities, m . The main aim will be to explore how the cost of supplying the service varies with m and to examine how changes in m may affect the consumer's cost and frequency of use. This is spatially equivalent to constructing a supply curve and examining its repercussions on users' price and demand.

Where the population has a fixed requirement for a particular service, an increase in the number of facilities reduces the time, cost and inconvenience of travel, provided they are sensibly located. Where demand is elastic, increased use has to be added to this list of direct benefits. These benefits have to be set against the costs of constructing and operating the extra facilities, normally paid through national and local taxes. Attempts will now be made in Chapters 2 and 3 to describe in turn how these costs and benefits vary as functions of m . A series of frameworks which enable the costs and benefits to be compared will then be outlined in Chapter 4.

Supply strategies

For a given budget, the provider of services faces a choice

between

- (1) a system with a few large facilities, and
- (2) a system with a larger number of small centres.

It may suit a particular region to combine these strategies by locating large facilities in concentrations of population and smaller ones in dispersed areas. In this section, however, the argument will concentrate on a context where all facilities have to be of the same size because the population is assumed to be uniformly distributed in order to simplify analysis. In subsequent sections the methods discussed for actually locating individual facilities do in fact allow combinations of large and small centres to occur in response to variations in density of population.

In manufacturing, economies can often be obtained through a larger scale of operation (Lloyd and Dicken, 1972). As Hirsch (1968) notes, relatively few attempts have been made to determine whether such economies of scale exist in the supply of public services. Other things being equal, the existence of economies of scale favours larger, more centralised facilities; conversely diseconomies of scale favour smaller, dispersed facilities. It is therefore important at the outset to ask what evidence there is for economies of scale.

Economies of scale

It is helpful to distinguish two ways in which economies of scale can occur in public services:

- (a) economies occurring through larger facilities being cheaper per unit of supply;
- (b) economies obtained by increasing the size of the units

of local government which manage the network of service centres, e.g., through spreading administrative costs.

Though possibly related, it is important to separate these effects. Only the first has a bearing on the present analysis. Unfortunately, most research on the topic has been stimulated by the issue of whether local governments should be consolidated into larger units and has therefore concentrated on asking whether larger authorities are more efficient. Although Massam (1975) does not make this distinction, virtually all of the studies he cites in reviewing work on economies of scale are of the second kind. Because of the emphasis on size of authority, unfortunately the distinction between effects of types (a) and (b) is sometimes blurred. Essentially, the difference is akin to that between economies of scale for single and multi-plant concerns in manufacturing industry (Lloyd and Dicken, 1972).

With few studies of the size of facilities available, inferences have to be made indirectly from studies of the size of authorities. To this end the relationship between size of authority and size of facility has to be clarified.

Suppose, not unrealistically, that the larger authorities of densely populated areas generally have larger libraries, fire stations, schools and other facilities. Where there are significant economies of scale in size of facility, the larger jurisdictions will reap these advantages. Thus, if analysis reveals that the larger authorities are more (or less) efficient, economies (or diseconomies) of type (a) might be inferred.

However, if larger authorities obtain economies in other ways, e.g., from spreading administrative and management costs over several

facilities, interpretation of these analyses is more complex. There is evidence that administrative costs in certain health services (Redcliffe-Maud Report, 1968) and education (Hirsch, 1968) increase less than proportionately with size of population up to a certain level (approximately 600,000 in the latter instance), but administration is generally a small proportion of total costs. Nevertheless, if we could assume under the heading of type (b) effects that larger authorities always enjoyed a net balance of economies over diseconomies, this would suggest that inferences based on size of authority would tend to overstate the advantages of a large size of facility.

In fact, straightforward inferences of this kind are often hard to make from existing studies because of the research design used. Most studies take a cross-section of authorities of different sizes and subject measures of performance and cost to statistical analysis. A curve relating average cost per head to size of population served is thereby fitted. A horizontal line indicates that average cost is independent of scale; a classical U-shaped curve suggests economies up to a certain level then diseconomies, the bottom of the trough giving the optimal size of facility or authority. Research comparing how well specific facilities of different sizes operate would of course be more relevant to the present discussion.

The problem with using a statistical cross-section is that in reality many authorities have a wide range of sizes of facility inherited from different periods in the past. Conurbations may inherit small facilities from peripheral authorities when their boundaries are extended. Thus diversity in size of facility within

authorities may lead to significant contrasts in size being partly lost in analysis carried out at the authority level. When assessing existing evidence, it must be borne in mind that the incidence of both economies and diseconomies of scale may thereby appear to be reduced, though not necessarily eliminated. This reasoning suggests that studies at the authority level may underestimate economies of facility size.

A further problem is that satisfactory measures of output are much harder to define, let alone obtain, than is the case with private goods and services. The task of measuring the output of hospitals, schools, libraries and refuse disposal services must surely daunt the heart of the most resolute econometrician. Although measures such as patient days in hospital, number of pupils taking S.C.E. or G.C.E. certificates, volume of books borrowed and miles of street cleaned may be forced into service for lack of anything better, the man in the street may justifiably wonder whether the qualitative benefits of competent medical care, a stimulating educational environment, a good library and a clean city somehow elude statistical detection. In fact, public opinion may be impervious to quantitative evidence: the number of pupils passing S.C.E. examinations has increased but there is currently much discussion of a decline in educational standards in secondary schools. In passing, it may be noted that the elusive qualitative benefits of many public services makes it hard to assess the competing claims of the private and public sectors for resources.

Two major sources of evidence are the research reports of the Royal Commission on Local Government in England and Wales (Redcliffe-Maud Report, 1968) and the reviews of research provided by Hirsch

(1968 and 1973). Both rely heavily on statistical analysis of cross sectional data. The Redcliffe-Maud Commission selected four services (housing, highways, certain health services and education) and correlated various measures of both scale of operation and volume of output for the four types of local authority then in operation. The results for the first three of these services are summarised in Table 2.1.

This summary table suggests that economies of scale are far less widespread than had been expected: economies are recorded for only 14 of the 38 entries. Furthermore, the graphs and equations presented in the report indicate that many economies actually identified were relatively small (Redcliffe-Maud Research Study 3, p.42). The fact that some of the best attested economies were for management and administrative costs (Research Study 3, p.21) suggests that such advantages as large authorities may be observed to have cannot be attributed solely to size of facilities. The evidence for economies is strongest in the case of highways; in housing reported diseconomies outnumber economies; for the health service studies the two kinds of entry are roughly equal. As measures of the quality of output were not included, it should not be forgotten that apparent diseconomies could actually reflect a better quality of service - a drawback of the design of this relatively unstructured study.

Research findings on a wider range of public services in N. America have been drawn together by Hirsch (1968); these broadly concur with those of the Maud Report. From this evidence (Table 2.2) only power supply and sewage treatment manifest continuous economies of scale; fire protection and school administration have

TABLE 2.1

Summary chart showing existence of economies (diseconomies) of scale

Service	Scale Variable	Dependent Variable	County Councils	County Boroughs	Non-County Boroughs	Urban District Councils	Rural District Councils
Housing	I. Population II. Total Stock of Dwellings	(a) Cost per foot super Supervision and Management	-	Diseconomies	Diseconomies	Zero Effect Economies	Diseconomies
		(b) Cost per foot super Supervision and Management	-	Zero Effect	Zero Effect Economies	Economies	Zero Effect
Highways	I. Population II. Total mileage	(a) Total Expenditure per mile	Economies	Economies	Economies	Diseconomies	-
		(b) Unclassified Road Expenditure per mile	Zero Effect	Zero Effect	Economies	Zero Effect	-
Health Service	I. Population II. Total Number of Visits	(a) Cost of Ambulance Service (per 1000)	Economies	Economies	Economies	Zero Effect	-
		(b) % Admin. costs Cost per Visit (Home nursing)	Zero Effect	Diseconomies	-	-	-
		(c) Cost per Visit (Home Nursing)	Zero Effect	Economies	-	-	-

Source: Redcliffe-Maud Research Report No. 3, 1969.

TABLE 2.2Cost curve studies of scale economies

Name and Year	Service	Type	Result
<u>Horizontally integrated services</u>			
Riew (1966)	Secondary education	S	AUC is U-shaped with trough at about 1700 pupils
Kiesling (1966)	Primary and secondary education	S	AUC is about horizontal
Hirsch (1959)	Primary and secondary education	S	AUC is about horizontal
Schmandt-Stephens (1960)	Police protection	S&Q	AUC is about horizontal
Hirsch (1960)	Police protection	S&Q	AUC is about horizontal
Will (1965)	Fire protection	E	AUC is declining with major economies reached at 300,000 population
Hirsch (1959)	Fire protection	S	AUC is U-shaped with trough at about 110,000 population
Hirsch (1965)	Refuse collection	S	AUC is about horizontal
<u>Circularly integrated services</u>			
Hirsch (1959)	School administration	S	AUC is U-shaped with trough at about 44,000 pupils
<u>Vertically integrated services</u>			
Nerlove (1961)	Electricity	S	AUC is declining
Isard-Coughlin (1957)	Sewage plants	S	AUC is declining
Lomax (1951)	Gas	S	AUC is declining
Johnston (1960)	Electricity	S	AUC is declining

Note: the following abbreviations are used:-

- S = statistical data
- AUC = average unit cost
- Q = questionnaire data
- E = engineering data

Source: Hirsch (1973)

U-shaped curves; education, police and refuse collection probably have horizontal average cost curves. It may be worth noting that the studies of education by Riew and Kiesling included qualitative measures such as the change in reading scores of children during a given time period.

Thus the majority of studies in both sources identified no economies of scale. With the exception of fire protection, it is striking that the services based on central facilities (education, police and the ambulance service) tended to have horizontal cost curves. If larger facilities are more efficient in any of these services, this higher efficiency is not great enough to manifest itself at the local authority level. It can still be argued of course that real economies arising from size of facility have been obscured through the diversity in size of facility found within authorities. Yet the difference in size between the smallest and largest county councils and county boroughs in England and Wales was substantial and may well have been great enough to allow size of facility to affect these results noticeably, if economies of scale did exist. The evidence therefore suggests that, over the main portion of the normal range of operation, economies of scale in services with central facilities are minor or non-existent.

The preceding inferences are somewhat indirect. Taking a more theoretical approach, we can ask whether the factors which generate economies of scale in manufacturing are likely to apply to public services. Lloyd and Dicken (1972) give three general reasons for economies of scale and a fourth may be added:

- (1) specialization of manpower and equipment;
- (2) economies of massed reserves;

- (3) economies of large-scale purchasing;
- (4) spreading of research and development costs over a greater volume of output.

From Lloyd and Dicken's discussion, the first would generally appear to be the most powerful.

The difference between public services (indeed, services generally) and manufacturing can now be seen quite clearly. With some exceptions (e.g., water supply and sewerage) public services are often labour intensive by their very nature: salaries and wages overshadow all other factor prices (Hirsch, 1973, p.309), often accounting for more than two-thirds of the current costs (Hirsch, 1968, p.509). With almost all its costs in labour, education exemplifies this clearly. Given the same average salary and size of class, a school for 1500 children will cost twice as much to run as a school for 750 children. It is therefore hardly surprising that the Redcliffe-Maud research report on education also found little evidence that larger authorities were more efficient.

Education is an extreme case of how the intrinsic face to face nature of a public service can govern the cost curve. Thus, where labour is the major cost and manpower requirements are proportional to the number of people served, economies of specialised machinery, specialised manpower and mass production can rarely be realised. The larger labour costs as a proportion, the flatter the average cost curve is likely to be. To varying degrees the library and fire service share this characteristic with education, although the specialisation of manpower possible in larger facilities may sometimes enhance efficiency.

Economies of massed reserves arise through larger plants needing to hold proportionately fewer spare parts or reserves of material and equipment. For reasons just discussed, these obviously apply far less to public services. Although economies of large-scale purchasing may apply to such inputs as power, Hirsch (1973, p.333) points out that counties and cities purchase a highly diversified array of factors and that few of these are bought in quantities large enough to obtain major price concessions. Even with lower prices for bulk purchasing, labour costs still predominate. Thus only about 20% of the library budget of the City of Edinburgh District Council is actually spent on books. If these economies are not likely to be important at the authority level, they can hardly be significant at the facility level. A similar conclusion applies to the spreading of research costs, though this is a very insignificant item anyway.

The advent of computerised record and information processing systems has probably created economies of scale in administration at authority (and possibly facility) level. In the services most affected (e.g., planning, social security, police and administration generally) this is more an argument for larger authorities than for larger facilities. In any case, as late as 1978 Coppock and Barritt observed that few authorities in Scotland were adequately equipped to exploit computers in one of the more promising areas for their application, namely planning. With the recent advent of micro-computers, which are relatively cheap, a smaller scale of operation may no longer be so disadvantageous in this respect.

Thus factors producing economies of scale are mostly absent both at facility level and at local authority level. With the



results of the empirical studies known, it is safe to say that they ought to have been expected on theoretical grounds.

Diseconomies of scale

Provision of a service in larger centralised units creates large working organisations. The general disadvantages of large organisations have been described persuasively by Schlesinger (1966):

"Large organisations suffer from a geometric increase in the difficulty of (a) successfully communicating intentions and procedures, (b) establishing a harmonious system of incentives, and (c) achieving adequate cohesion among numerous individuals and sub-units with sharply conflicting wills ... Large organisations find it hard to anticipate, to recognise, or to adjust to change."

Such disadvantages may help to explain the rising portion of the U-shaped curves noted by Hirsch for fire services and school administration.

These attributes of a larger scale of operation may be important, however intangible in cost terms, since they may colour the attitude of employees and the experience of users in a way which lowers the quality of service. Moreover, in services where a high participation by the public is desirable to promote democratic control, there may be serious disadvantages in centralising the service into a few large units, whether these are more efficient or not. What weight this argument carries in relation to a particular service will of course depend partly on the individual's system of values and his political philosophy.

Schlesinger's arguments are intrinsically hard to substantiate with firm evidence, though observed diseconomies of scale lend

indirect support. At present there is pervasive anecdotal evidence on the disadvantages of the large secondary schools (normally around 1500 pupils) introduced in many areas of Scotland in the last decade. The comments of teachers with experience of smaller schools frequently echo Schlesinger's remarks about the anonymity, inflexibility and lack of communication at the larger scale.

The disadvantages of large units are taken very seriously in Norwegian education. In 1969 the Norwegian Parliament passed an education act which stated that, normally, a secondary school should not have more than 450 pupils. As a result the mean size of secondary school in Norway has changed very little in recent years, increasing very slightly from 311 in 1968-69 to 322.5 in 1973-74 (Statistisk Sentralbyrå, 1969 and 1975). These figures present a striking contrast with Scotland where the mean size of secondary school almost doubled in a similar period, increasing from 499 in 1969 to 862 in 1974 (HMSO, 1969 and 1977). The fact that the qualitative disadvantages of a larger scale are sufficiently tangible in at least one society to precipitate legislation would seem to indicate a need for more thought and research on this topic.

Two further pieces of evidence might be adduced. First, Volvo has reorganised car assembly in the hope of replacing the boredom and anonymity of mass production with the personal satisfaction and flexibility of a small working environment. Second, a fascinating but unreplicated study of a large sample of six year olds in rural Dorset (Lee, 1957) shows that longer journeys to school, particularly by bus, adversely affect the children's "adjustment" to school and their enjoyment of it. The continuing consolidation of rural schools into larger units implies longer journeys and may therefore carry the

social cost identified by Lee, which can be viewed as a disadvantage of size, albeit an indirect one.

This evidence, tangential though it may be, suggests that research may have been overconcerned with the frequently non-existent economies of a larger scale and insufficiently sensitive to the attractions, often qualitative, of a smaller scale, including better accessibility. Such attractions, however, are not purely qualitative as the numerous diseconomies shown in Tables 2.1 and 2.2 testify. Nevertheless, in a local context it is doubtful whether the local libraries, swimming pools or fire stations in Edinburgh, excepting possibly the one or two largest, are big enough to suffer seriously from either disadvantage.

Conclusions on effects of scale

Hirsch (1968, p.517) summarised the state of knowledge thus:

" ... it appears that the following urban government services are likely to enjoy major economies of scale: air pollution control, sewage disposal, public transportation, power, water, public health services, hospitals and planning.

Most of the other urban government services for government units of more than about 50,000 inhabitants are likely to enjoy only minor, if any, economies of scale.

This does not deny that certain specialized higher education and library facilities can incur scale economies; but these appear to be the exception when compared to the major education and library expenditures."

Later Hirsch (1973, p.332) said that economies of scale were uncertain for hospitals and very minor for fire services. Unfortunately this summary does not distinguish between the effect of size of facility

and size of authority and makes no mention of diseconomies.

In conclusion, at present the most reasonable assumption for central facilities over typical variations in operating scale is that capital and operating costs increase linearly with size, i.e., the curve for average unit cost is horizontal. Cost curve (f) rather than (e) or (g) (Fig. 2.1) is therefore adopted as the most plausible basis for further analysis, but the others are not excluded. Another way of looking at curve (f) is to assume that only a standard size of facility, with one fire engine say, is feasible and that facilities come in multiples of this. Hence (f) indicates that stations with two or three engines cost twice and three times as much as a station with one engine. For libraries the equivalent assumption is that the size of library which two librarians can manage efficiently will be twice as costly as the size one librarian can look after efficiently and so on. The size of swimming pools conveniently tends to be fixed by requirements for competitions but engineering data may also have a bearing on size and cost in this case.

The form of the cost curve has a significant bearing on the number of facilities which should theoretically be provided. Before this can be explored, the relationship between the number of facilities, m , and consumers' cost and frequency of use must be discussed.

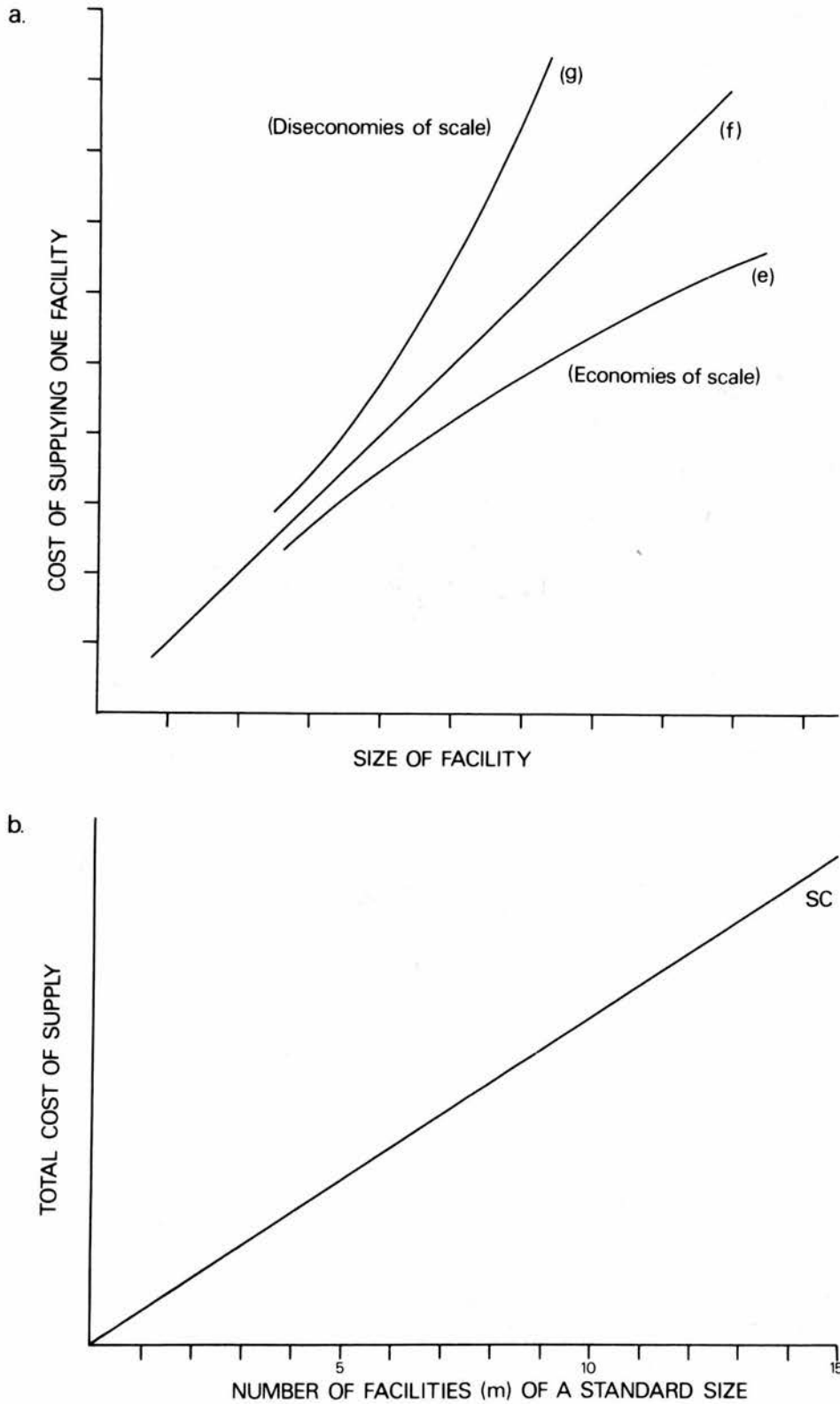


Figure 2.1

Assumed relationship between:

- (a) size of facility and cost of supplying one facility
- (b) number of facilities (m) and total cost of supply (sc)

CHAPTER 3

The benefit to users of better access to facilities

Aim

The preceding chapter was concerned with supply; discussion centred on the form of the relationship between the cost of providing the service and the number of facilities, m . The present chapter will be concerned with demand and will examine various ways in which users benefit from better access to facilities. The question of finding specific locations for a given number of facilities will be examined in the next section.

To be more specific, we wish to answer three main questions in this chapter:

- (a) given fixed demand, how does aggregate travel fall as the number of central places increases?;
- (b) given elastic demand, how does the volume of use or demand rise as m increases?;
- (c) can we find a more general measure of social benefit which takes account of both travel cost and level of use?

As questions which can be asked about a system of central places, these are quite fundamental. Despite their importance, it would appear that they have received very little explicit treatment in the classical work of Christaller (1966) and Lösch (1954) or in the more extensive, recent work on systems of central places, e.g., Berry (1967), Marshall (1969).

The method devised to answer these questions involves invoking some of the traditional assumptions of location theory: a uniform distribution of population or demand; hexagonal catchment areas; and

an isotropic plane with straight-line journeys to the central point. Initially we also assume that consumers use the nearest centre. Later it will be shown that some of these assumptions, particularly that relating to population density, can be relaxed without significantly affecting the conclusions reached.

Size of catchment areas as m increases

The first step is to determine how the size of catchment areas will change as m increases. Some attempts were initially made to create a model showing how catchments alter in size for the bounded space within a city, but these were unsuccessful. It therefore proved necessary to assume an unbounded plane and to use firstly, a Christaller landscape, and then a somewhat more flexible system of catchments akin to that postulated initially by Lösch.

Of Christaller's three models of central place hierarchies, the system based on the 'marketing principle' (with $k=3$) seems most suitable for the present purposes, since travel by consumers has a lower total value than in the models based on $k=4$ and $k=7$. All of Christaller's landscapes have a rigid hierarchical ordering such that the number of places increases in a fixed multiple from a given level to the one below. For $k=3$ this arrangement requires that the number of facilities is either 1 or 3 or 9 or 27 etc. but cannot be 2 or 6 or 10. Essentially, this results from the underlying principle that successive lower orders are located in the centre of the interstices between centres of the level in question. Such a restriction on m is completely unrealistic as a limitation on the number of facilities required for an urban area. At present, however, we merely wish to look at sizes of catchments and the Christaller landscape provides a convenient departure point for doing so.

Consider a given hexagonal catchment area around a facility, F (Fig.3.1a). The circle circumscribed around the hexagon has radius denoted by 'r max' which we will refer to as the maximum radius of the hexagon. The distance 'a' from F to the midpoint of a side of the hexagon is also the radius of the inscribed circle. Here a will be called 'the minimum radius of the hexagon' or 'rmin'.

The area of the hexagon will be six times the area of each equilateral triangle composing it or twelve times the area of the shaded triangle, DEF, in Figure 3.1a . Since the area of a triangle is given by $\frac{1}{2}$ (height).(base), it is clear from Figure 3.1b that

$$\text{area of } \triangle DEF = \frac{1}{2}a \cdot \frac{1}{2} r_{\max}.$$

Since

$$\tan 30^\circ = \frac{\frac{1}{2} \cdot r_{\max}}{a},$$

$$\frac{1}{2} r_{\max} = a \tan 30^\circ = \frac{a}{\sqrt{3}}.$$

Hence

$$\text{area of } \triangle DEF = \frac{1}{2} \cdot a \cdot \frac{a}{\sqrt{3}} = \frac{a^2}{2\sqrt{3}}.$$

Thus

$$\text{area of hexagon} = \frac{12a^2}{2\sqrt{3}} = 2\sqrt{3} a^2$$

The area of a hexagonal catchment area can therefore be expressed in terms of the radius, a, of its inscribed circle. It is convenient to note for later reference that rmax can be expressed as $(2/\sqrt{3})a$ or $(2/\sqrt{3})r_{\min}$.

Let $a_1, a_3, a_9, \dots, a_m$ be the minimum radii of the 1,3,9... m regions tributary to the 1,3,9 ... m centres of the 1st, 2nd 3rd ... n^{th} order of a Christaller landscape with $k=3$. Though the plane is treated as unbounded, the network of hexagons associated with each order covers its whole area in effect; the total area covered by each order of catchments is therefore the same. Since there are three times the number of hexagons in the second order as the first, the area covered by a hexagon of the first order will be

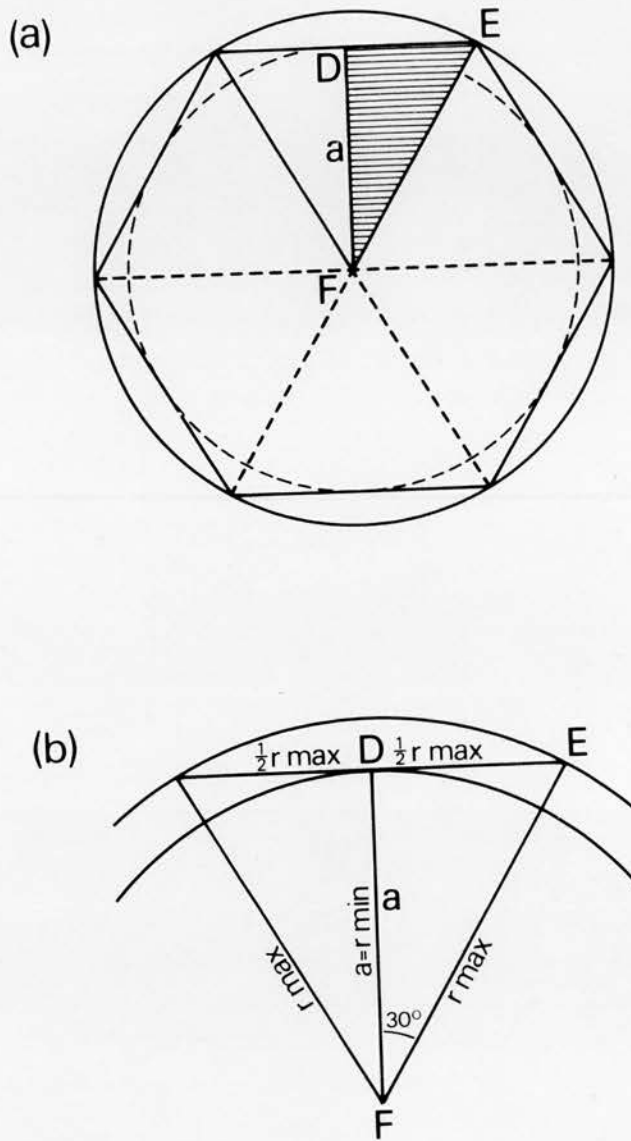


Figure 3.1

Hexagonal catchment area with inscribed and circumscribed circles

three times that covered by a hexagon in the second order. In short

$$\begin{aligned}
 2\sqrt{3} a_1^2 &= 3(2\sqrt{3} a_3^2) \\
 \text{i.e. } a_1^2 &= 3 a_3^2 \\
 \text{i.e. } a_3 &= \frac{a_1}{\sqrt{3}} = a_1 \cdot 3^{-\frac{1}{2}}
 \end{aligned}$$

More generally, for the m centres¹ of the n^{th} order

$$\begin{aligned}
 2\sqrt{3} a_1^2 &= m 2\sqrt{3} a_m^2 \\
 \text{i.e. } a_1^2 &= m \cdot a_m^2 \\
 \text{i.e. } a_m &= \frac{a_1}{\sqrt{m}} = a_1 m^{-\frac{1}{2}}
 \end{aligned}$$

Thus, when there are m centres each has a tributary region whose minimum radius can be expressed in terms of m and a , which is a constant of the system.

Christaller's principle of interstitial location requires that a centre in any order will also function as one of the centres in all orders below. For instance, in siting nine facilities, three would be placed automatically at the locations occupied when only three centres were in existence. Christaller's models therefore incorporate a very strong hierarchical constraint on location. At a regional scale this principle has much validity: in an area of dispersed population an existing town or village has considerable inertia, derived from its infrastructure and such factors as local shopping habits. Under these conditions it makes sense to keep the number of new centres down to a minimum, as Christaller's principle does in essence. However, within

1. where m is such that $m = 3^{n-1}$. In a $k=4$ landscape $m = 4^{n-1}$; in a $k=7$ landscape $m = 7^{n-1}$. Since the generalisation derived subsequently does not depend on the specific relation between m and n , it is true for all three Christaller landscapes.

a city a facility can be potentially located anywhere in the street network, subject to site requirements; it therefore seems preferable to dispense with the hierarchical constraint here, since we are more interested in the latter context.

It is worth noting that a satisfactory dynamic version of central place theory does not exist at present and a good conceptual framework for treating the sequencing of centres is hence not available. As a result the time dimension within which these decisions are made cannot be incorporated into the present framework. Despite this, the locational algorithms discussed in the next section do provide a useful means of exploring possible sequences in any specific context.

Abandoning the hierarchical constraint on location, imagine a landscape in which any number of facilities can exist. The hexagonal tributary areas are now allowed to vary almost continuously in size (Fig.3.2); the simple hierarchy of Christaller no longer exists. This landscape is akin to that postulated by Lösch before the 'rotation' stage in his model of urban places.

If the preceding argument is extended to the new landscape, the 2,3,4... m centres serve the same total area as the first centre. Hence

$$2\sqrt{3} a_1^2 = m 2\sqrt{3} a_m^2 ;$$

and therefore $a_m = \frac{a_1}{\sqrt{m}} = a_1 m^{-\frac{1}{2}} .$ (3.1)

We now have a means of determining the size of catchment areas (a_m), given the number of centres, m: here m can have any integer value.

Aggregate and mean travel within one catchment area, given inelastic demand

The next step is to derive an expression for aggregate travel (AT) to the centre from within one whole catchment area. For a

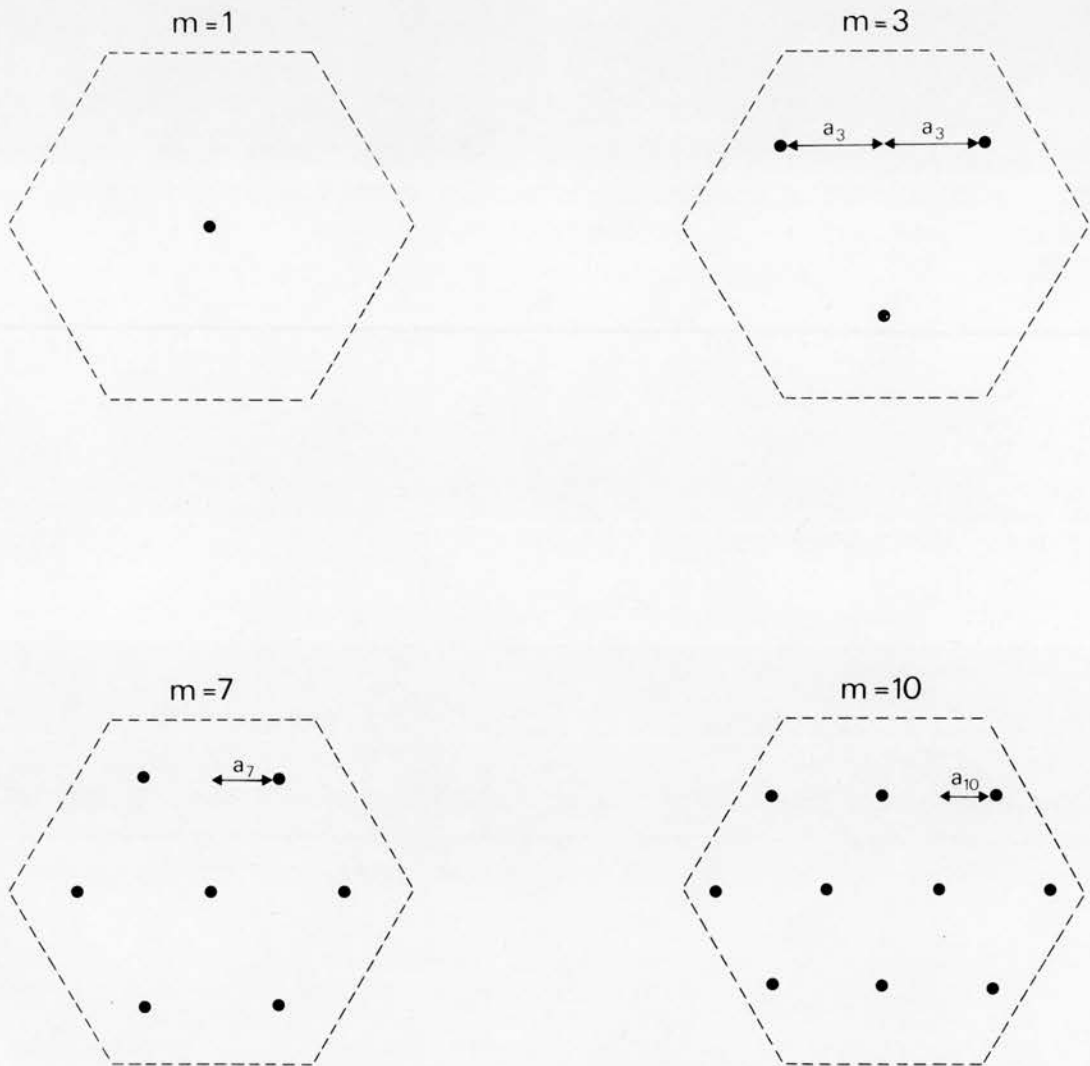


Figure 3.2

Landscape in which the number of central points
can have any integer value

circular or hexagonal catchment area we can do this by using integral calculus. Since the exposition is considerably easier for a circular catchment and since the approximation involved in treating a hexagon as a circle can be shown to have little effect on the result, the catchment areas will be mostly treated as circles for the rest of this chapter. The inscribed circle gives a closer approximation to the area of the hexagon than the circumscribed circle and will be used in preference to the latter.

Consider a circular catchment area of radius a containing an annulus or ring, denoted by i , which is r units from the centre, j , and has width δr , where δr is infinitesimally small (Fig.3.3). If the annulus were unwound into a long straight strip it would be approximately rectangular, with width δr and length $2\pi r$ (the inner perimeter of the ring). Hence the area of the annulus is approximately $2\pi r \cdot \delta r$ and, if the density of population in the catchment is p , the total population of the ring is $2\pi r \cdot \delta r \cdot p$. If each member of the ring's population makes one trip (of distance r) to the centre, then total travel from the ring, denoted by AT_i , is given by

$$AT_i = 2\pi r \delta r p r = 2\pi r^2 p \delta r$$

We can think of the whole circle as composed of a host of infinitely slender annuli like the first one. Each has width δr ; the smallest adjoins the centre; the largest is bounded on the outside by the perimeter of the circle at a distance of a units from the centre. Now if each member of the population makes one trip to the centre no matter how far away, aggregate travel from the whole catchment, AT_j , will then be the sum of the distances travelled from all of these rings. It can therefore be found by integrating

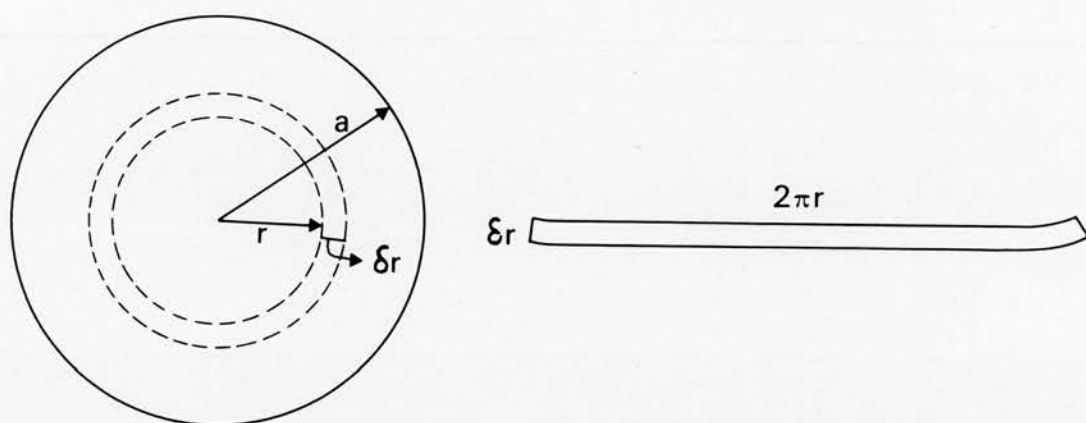


Figure 3.3

Unwinding a ring within a circular catchment to form a rectangular strip

the above expression with respect to r between the limits 0 and a , (the distance of the shortest and longest trips). Thus aggregate total travel within the whole catchment is

$$\begin{aligned}
 AT_j &= \int_0^a 2\pi r^2 \cdot p \cdot dr \\
 &= 2\pi p \int_0^a r^2 \cdot dr \\
 &= 2\pi p \left[\frac{1}{3} r^3 \right]_0^a \\
 \text{i.e. } AT_j &= \frac{2}{3} \pi p a^3 = \frac{k_1}{3} a^3 \\
 &\quad \text{where } k_1 \text{ is } 2\pi p
 \end{aligned}$$

Similarly mean travel for the whole catchment, MT_j , is given by

$$\begin{aligned}
 MT_j &= \frac{\sum \text{population of each ring} \times \text{distance to centre}}{\sum \text{population of each ring}} \\
 &= \frac{\int_0^a 2\pi r \cdot p \cdot dr}{\int_0^a 2\pi r \cdot p \cdot dr} = \frac{2\pi p \int_0^a r^2 \cdot dr}{2\pi p \int_0^a r \cdot dr} \\
 &= \frac{\left[\frac{1}{3} r^3 \right]_0^a}{\left[\frac{1}{2} r^2 \right]_0^a} = \frac{2}{3} a
 \end{aligned}$$

Hence the mean distance travelled within a circular catchment area is simply two thirds of the radius. It may also be worth noting that the expression for total travel, AT_j , is simply mean distance multiplied by the population of the circle, $\pi a^2 \cdot p$. Since we have been assuming that the number of trips to the centre does not vary with distance it is worth emphasizing that the argument has been based entirely on inelastic demand so far.

Aggregate and mean travel for the whole system with inelastic demand

In the landscape postulated above, mean travel will be the same in each catchment, no matter what value m has. Hence if mean travel can be derived for one of the catchments, this will give mean travel for the whole system. Now the total population of the whole area can be taken as the population of the largest catchment (defined by $m=1$) which is $\pi a_1^2 p$, when treated as a circle or,

more accurately, $2\sqrt{3} a_1^2 p$, if treated as a hexagon. In either case as m increases and catchments become smaller, total population of course remains constant. For any given value of m total travel in the system will be given by mean travel from one catchment times total population, i.e., total travel can be regarded as a simple constant multiple of a variable mean travel. It follows that a curve depicting the relationship between m and aggregate travel will have the same shape as one for mean travel and m . Since the mathematical expression for MT is simpler; it will generally be used below in preference to AT.

It has already been established that the minimum radius of a set of m hexagonal catchments is $a_m = a_1 m^{-\frac{1}{2}}$ (3.1); this expression represents the radius of the circle which can be inscribed within each hexagon. Now mean travel from any circular catchment is two thirds of its radius; for the circular catchments, therefore,

$$MT = \frac{2}{3} a_1 m^{-\frac{1}{2}} \quad (3.2)$$

This argument uses the inscribed circle to approximate the hexagon.

Since it can be shown that MT from a hexagonal catchment¹ is $.7020 a_1$, MT can be expressed more precisely as $.7020 a_1 m^{-\frac{1}{2}}$.

Now aggregate travel in one hexagonal catchment of minimum radius a_m will be mean travel times the population of that catchment; hence

$$\begin{aligned} AT_j &= .7020 a_1 m^{-\frac{1}{2}} \cdot 2\sqrt{3} a_m^2 p \\ &= .7020 a_1 m^{-\frac{1}{2}} 2\sqrt{3} a_1^2 m^{-1} p \quad (\text{using 3.1}) \\ &= (.7020) 2\sqrt{3} a_1^3 m^{-\frac{1}{2}} p . \end{aligned}$$

1. I am indebted to John Martin for carrying out the complex integration required to obtain mean travel for a hexagon. The use made of calculus beyond this point and any errors in it are my own responsibility.

To obtain aggregate travel for all m catchments in the landscape,

$\sum_j AT_j$, we multiply by m :

$$\begin{aligned}\sum_j AT_j &= m(.7020) 2\sqrt{3} a_1^3 m^{-1/2} p \\ &= (.7020) 2\sqrt{3} a_1^3 p m^{-1/2} \\ &= k_{AT} m^{-1/2}\end{aligned}\quad (3.3)$$

$$\text{where } k_{AT} = (.7020) 2\sqrt{3} a_1^3 p$$

Despite superficial differences the essential form of relationships 3.2 and 3.3 is the same: mean and therefore aggregate travel is a simple inverse square root function of the number of centres.

The form of the relationship between aggregate travel and the number of central places in a landscape has now been established. By choosing any convenient arbitrary value for the constants, k_{AT} and a_1 , it is possible to plot the decline in MT and AT as m increases (Fig.3.4). Each successive centre contributes progressively less in terms of the absolute reduction in mean travel and the curve progressively flattens out above a certain level of m . For a Christaller landscape the relationship has the same general form but can only be evaluated when m is a multiple of three.

Aggregate travel in Edinburgh

It is interesting to compare this theoretical curve with the equivalent curve for the population of Edinburgh in 1971. To do this for a given value of m we select a set of efficient locations for these facilities. We then assume that each member of the city's population makes one trip to his or her nearest facility and compute the total travel involved. The locations are chosen by using a searching procedure incorporated in the computer program NORLOC (Törnqvist et al., 1971). Though not necessarily always identifying completely optimal locations, this algorithm gives relatively efficient sites;

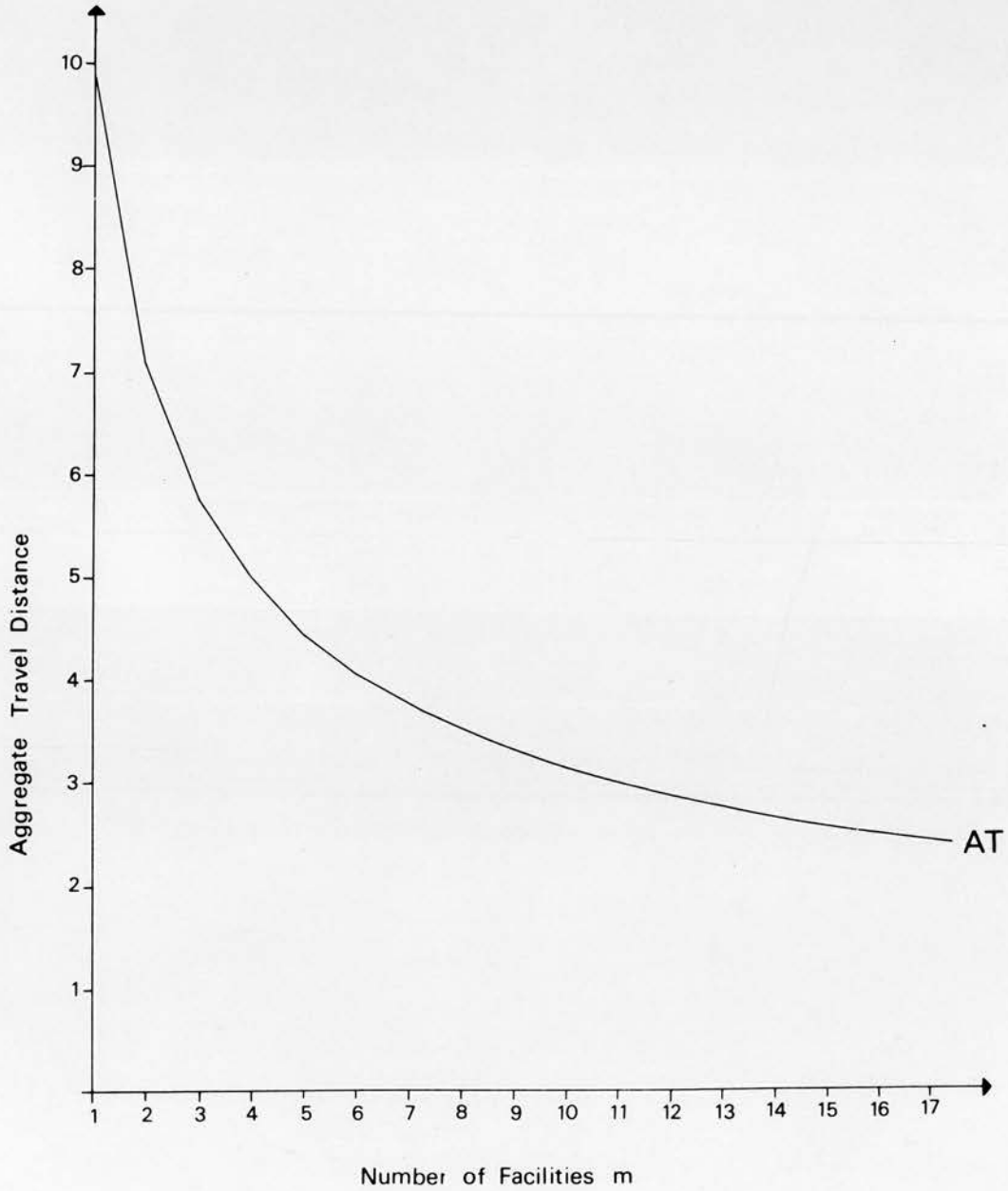


Figure 3.4

Relation between aggregate travel and m on an isotropic plane when demand is inelastic

discussion of its efficiency will be postponed to the next section, which is concerned with the purely locational aspects of the problem. To use the NORLOC program, population had to be allocated to cells in a square grid. A computer program (CONVERTS), which allocated each of 1346 census enumeration districts for the City of Edinburgh in 1971 to a framework of 500 metre cells and which satisfied the rather unusual input requirements of NORLOC was therefore written by the author. A modified version of CONVERTS, called PUNMAP4, was also written to present the census data in a grid form suitable for submission to the square grid mapping program, CAMGRID, and to another mapping program called GIMMS .

The distribution of Edinburgh's population shows striking variations in density (Fig.3.5). There are several extensive areas of relatively high density near the edge of the city, invariably council housing estates, which contrast with the lower densities of nearby private estates. Although density in the inner areas, including Leith, is generally high, Edinburgh does not conform well to the pattern of exponential decay in density from the centre typical of N. American cities. For instance, the significant contrasts in density on the periphery mean that density is more even in some sectors than a uniform decay would allow. Further, intermediate areas, which were built up in the interwar or early postwar period when the 'garden city' style was popular with urban designers, sometimes have lower density than the adjoining council estates further out.

Since the shape of the aggregate travel curve in a city will depend on the distribution of population, it is surprising at first that the curve for Edinburgh is very similar in form to that derived

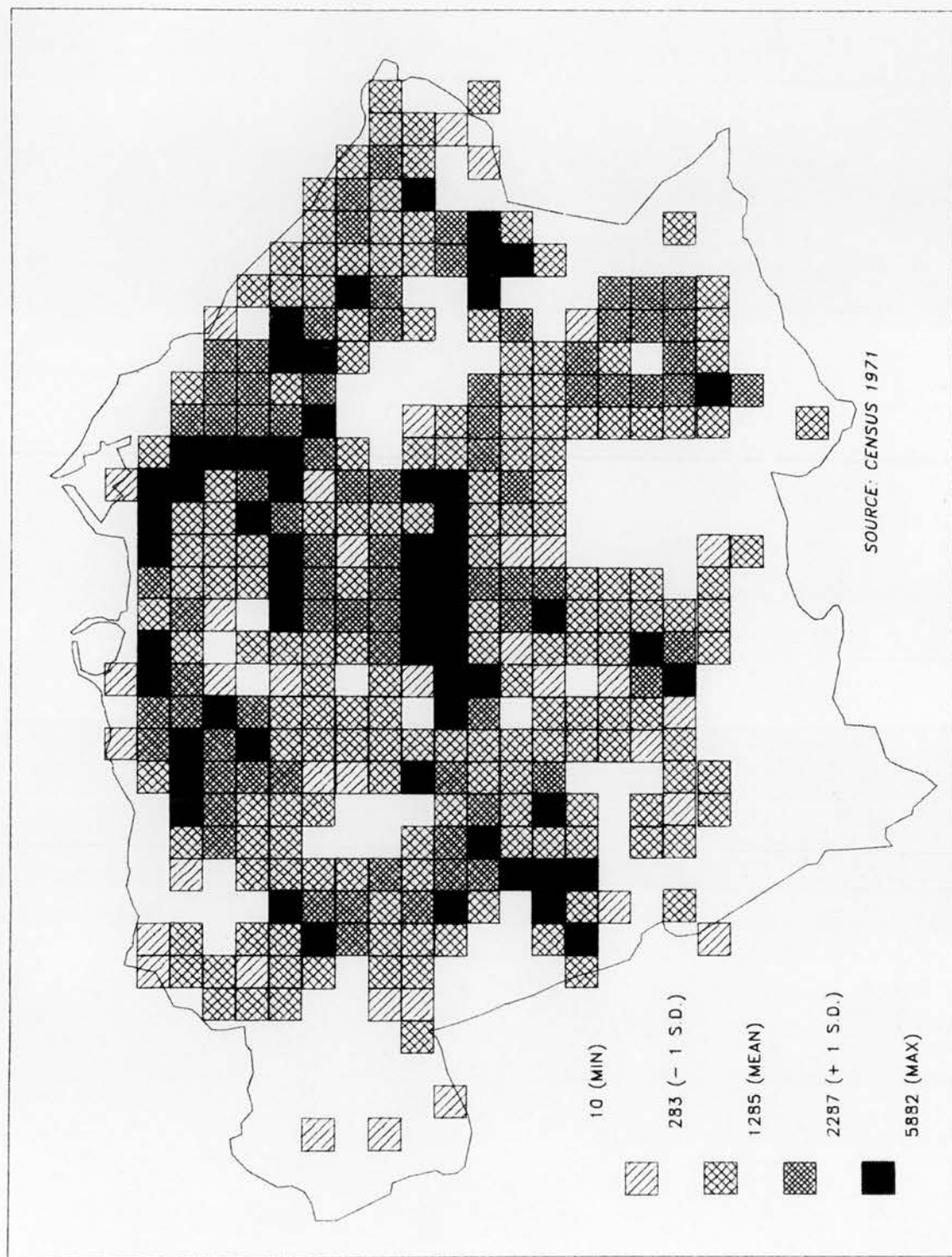


Figure 3.5 The distribution of population in Edinburgh by grid squares of 500 m in 1971

already on the assumption of uniform density (Fig.3.6). Clearly the correspondence between the two curves cannot be explained by a relatively even density in the city. Rather, the inherent strength of the underlying relationship must be great enough to offset the blurring caused by wide variations in density. Accordingly, it must be concluded that the general form of this curve is a relatively robust spatial property of urban systems.

Use of NORLOC entails the assumption that consumers travel in straight lines to the nearest centre. If we assumed instead that actual trips generally exceed straight line journeys by a certain factor, say 30% , the curve would start off at a higher point but retain the same shape. Similarly, an assumption that a consistent proportion of the population in any catchment did not use the nearest facility, would probably have a similar effect. However, if in reality a facility in the centre of town was predominantly used over the whole city in preference to outlying centres, this could alter the shape of the curve.

Elastic demand

The preceding discussion has assumed that demand is the same, no matter how many centres are built. In contrast, when demand is elastic, an increase in the number of centres will stimulate increased use as well as reducing mean travel cost. The amount of extra demand created will then be a measure of how valuable extra facilities are. By extending the method already employed, we can try to define a relationship between m and the level of use. Again we assume a uniform landscape and approximate hexagonal catchments by circles.

As noted in chapter one, it is important not to confuse the assumption of elastic demand with the well known distance decay effect around individual centres. This latter effect may describe a

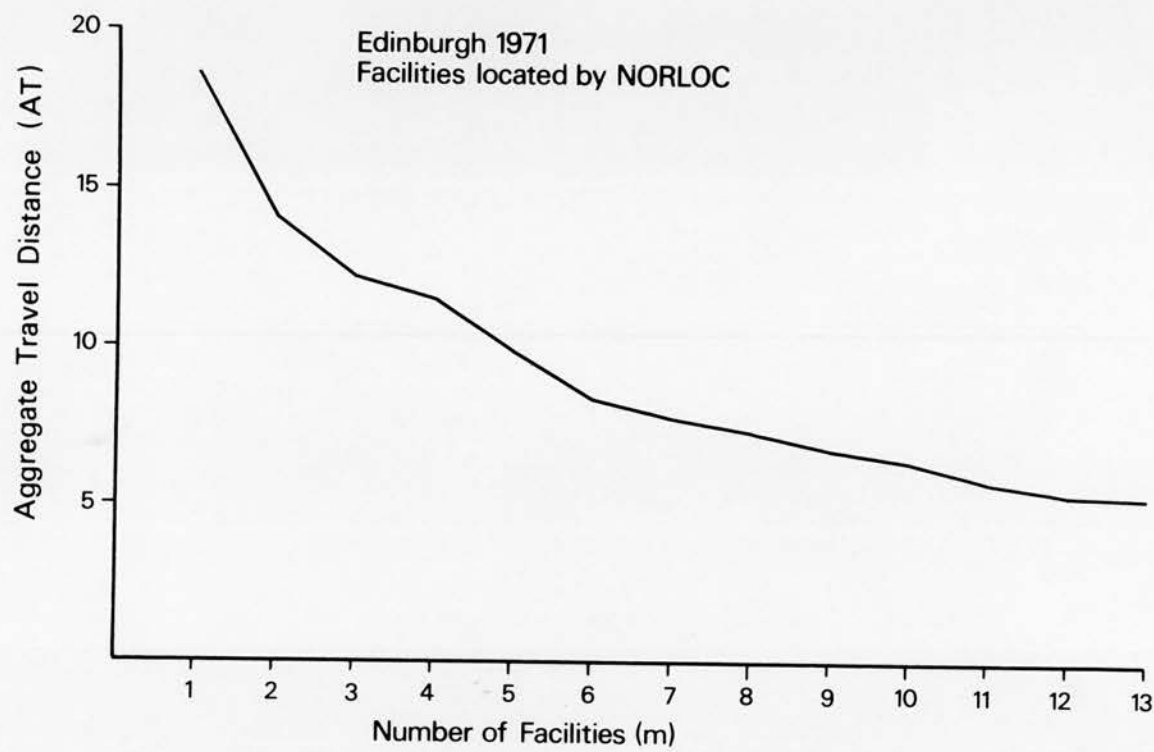


Figure 3.6 Relation between aggregate travel and m for
Edinburgh with centres located by NORLOC

genuinely elastic demand or it may describe the way in which distance affects consumers' choice of centre when the overall level of demand is nevertheless fixed. Or it may describe a mixture of the two. A simple example illustrating this difference is provided elsewhere (Hodgart 1978, pp21-22). In this context, it should be noted that spatial interaction models incorporating a distance decay of the type

$$T_{ij} = \frac{P_i}{d_{ij}^b} \quad \text{where } T_{ij} \text{ is the number of trips from cell } i \text{ to centre } j \text{ and } b \text{ is an exponent}$$

can in fact be used to describe travel patterns with elastic or inelastic demand depending on how they are constructed.

If we assume that catchment areas do not overlap, then all demand to a centre will come from within its surrounding hexagon, i.e. our catchments are 'deterministic'. Let f denote the number of trips to a centre per head of population (or per household) from a given area during a specified period. If f declines consistently with distance so that

$$f = \frac{k_2}{d^b} \quad (3.4.1)$$

where k_2 is a constant,

we have defined a framework in which demand is elastic. This is probably the simplest and most common formulation of elastic demand used in geography. Since it is so frequently used, it is worth exploring to see what conclusions it produces.

The total number of trips from an annulus, i , r units distant from the centre will be the product of its population, $2\pi r.dr.p$. and its frequency of trips, f . Hence

$$T_i = f \cdot 2\pi r.dr.p = \frac{k_2}{r^b} 2\pi r.dr.p \quad (3.4.2)$$

where T_i is the number of trips from ring i .

Following Lüscher (1954), we can regard the catchment areas now defined as cones of demand, sloping downwards in all directions from the central point. If the crucial parameter, b , (sometimes called the beta value) is large, then the frequency of trips will fall sharply away from the centre, giving a steep narrow cone (Fig.3.7a). On the other hand a low value of b gives a gentle decline in demand with proportionately more users coming from the edge of the catchment, thereby producing a broader less compact cone (Fig.3.7b). When $b=0$ there is no decline; instead a flat plateau of demand extends to the catchment's boundary; demand is then inelastic. Though Lüscher noted that the total demand could be obtained by integrating to find the volume of the cone and other writers have spoken of 'unwrapping' demand cones (Taylor, 1975), judging by the published literature no one has actually tried to put this possibility into practice or use it for further analysis.

In fact the distinction between elastic and inelastic demand has very rarely been made in empirical work on catchment areas. As a result little is known about which services have elastic demand, let alone how elastic. Although empirical studies in spatial interaction find values for b ranging from $\frac{1}{2}$ (a gentle decline) to around 2 (a steep gradient), these estimates are almost invariably made in a framework which is too loose to permit any distinction between elastic and inelastic demand and therefore provide little guide to empirical or theoretical work on the present lines.

Thus an introductory monograph on distance decay curves by Taylor (1975) provides a most valuable exploration of the mathematical form of distance decay curves, but contains no discussion of whether these curves represent elastic or inelastic demand. In view of the preceding discussion it may appear very surprising that such a

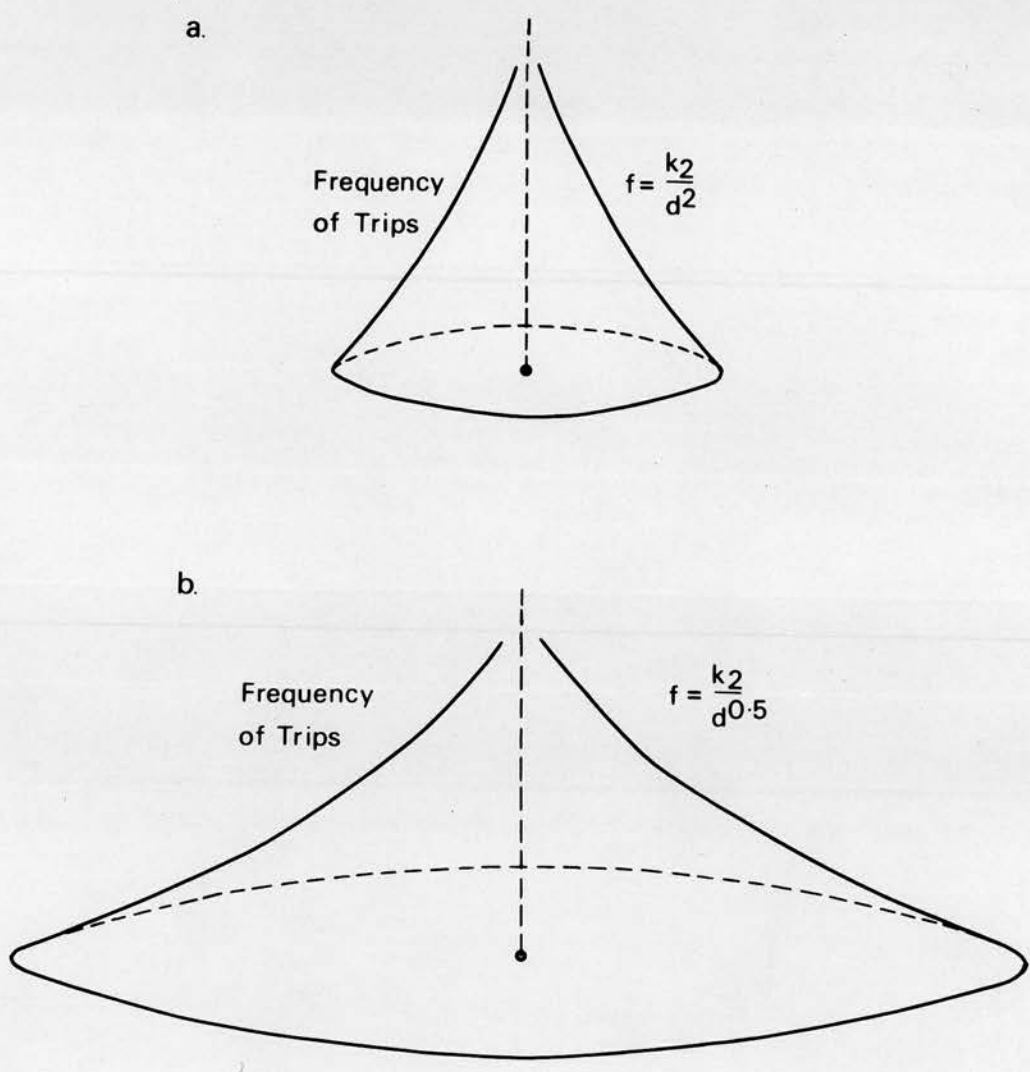


Figure 3.7 Demand cones defined by inverse powers of distance

distinction has not become central to location/allocation modelling. However obvious this point may seem now, the present research was well under way before the author became fully aware of the point himself. In fact, the point practically forced itself into the author's attention through repeated difficulties encountered in trying to write a computer algorithm for locating facilities, which would improve on Törnqvist's method.

Part of that algorithm attempted to use population potential within the immediate vicinity of a possible supply site as a rough guide in selecting starting points for locational search. From this exercise it was realised that what population potential really meant in terms of an underlying rationale for locating a facility was not properly understood. Eventually this train of thought led to the realisations that population potential could be taken to represent a situation where frequency of use fell with distance from the centre and that the algorithms of Törnqvist and Cooper were based on the important, yet unstated, assumption of inelastic demand. Curiously the latter step came first and the former some time after. The algorithm then under construction was eventually abandoned as too clumsy to be very useful. Though the method proved fairly impracticable, the writing of the program, nevertheless, yielded interesting theoretical insights.

There does not seem to be any work on population potential which makes the inherent connection with elastic demand. The concept of a spatially elastic demand was stated explicitly by Lösch (1954), though it was then already implicit in Christaller's (1966) work on central places. Since then the idea of a spatially elastic demand has been neglected in studies of spatial interaction and virtually ignored in work on location/allocation algorithms. Accordingly, a major part of

the present work is concerned with examining the implications of elastic demand. An attempt is also made to apply this as an extra criterion in a limited context to a study of the locations of swimming pools.

Existing empirical studies give little guide to what values of b would be appropriate to use in examining the implications of elastic demand; nor does the literature provide much advice on what mathematical formulations would be most helpful. In fact it is interesting that a study in which the catchments of swimming pools are simply portrayed clearly in map form (Currie, 1977) provides some useful evidence suggesting that distant users make less use of pools than people with better access. Surprisingly, studies based on more complex statistical approaches have often not been able to do so.

Part of the reason is that distance decay curves cited in the literature are based on observations of visits to one centre or a series of centres; they do not compare the frequency of use of households at different distances. It is therefore impossible for the former to disentangle the extent to which such curves represent a fall in overall use of the service or merely a fall in preference for the specific centre. On common sense grounds one would expect that, for the most essential services such as medical care, they mostly represent the latter effect. For less immediately essential services, such as recreation, these curves probably encapsulate a significant amount of both effects. In such cases the fall in overall demand will be less steep than the purely descriptive decay curve itself, since the total "decay" effect only arises in part from the fall in demand. We can then say that the value describing the degree of elasticity will be less than that for the descriptive curve and is therefore likely to be less than two, since most b values obtained in calibrating gravity

models are less than that figure, which does not really narrow it down very much, however.

Faced with a lack of information on the main parameter, b , the best strategy seems to be to take a wide enough range of values to cover all the likely cases. Thus if we take the range from $b=0.25$ to $b=2.00$, we can be fairly sure that most services where demand is significantly elastic will lie somewhere between these extremes.

An expression has already been presented for the numbers of trips from one annulus, (3.4.2). Demand from the whole catchment, T_j , can be obtained by integrating this expression between the limits 0 and a . Hence

$$\begin{aligned} T_j &= \int_0^a k_2 \cdot 2\pi p \cdot r \cdot dr / r^b \\ &= k_2 k_1 \int_0^a r^{1-b} \cdot dr \quad (k = 2\pi p) \\ &= k_3 \int_0^a r^{1-b} \cdot dr \quad (3.5) \end{aligned}$$

$$\text{where } k_3 = k_2 k_1 = k_2 2\pi p \quad (3.5.1)$$

In general this integration is straightforward since

$$\int n \cdot x^{n-1} \cdot dx = x^n + c \quad (3.6)$$

However, the case when $n=0$ is an exception to this rule and the following relationship has to be used:

$$\int x^{-1} \cdot dx = \log_e x + c \quad (3.7)$$

(Wilson and Kirby, 1975, p.104)

In our problem the exceptional case occurs when $b=2$, but for all other values of b , (3.6) can be applied.

The case when $b=1$ yields a result of particular interest:

$$\begin{aligned} T_j &= k_3 \int_0^a r^{1-b} \cdot dr = k_3 \int_0^a r^0 \cdot dr \\ &= k_3 [r^1]_0^a = k_3 a \quad (3.8) \end{aligned}$$

Thus the number of trips from a circular catchment is directly proportional to the radius of the catchment; doubling the radius of the catchment, simply doubles the number of trips. Now, if we divide the catchment into a series of concentric rings each one unit wide, the total demand from any ring is given by evaluating the integral in (3.8) between the appropriate limits. Hence if the number of trips from the third ring, lying between a distance of 2 and 3 units from the centre, is denoted by T_3 then

$$\begin{aligned} T_3 &= k_3 [r^1]_2^3 = k_3 [3 - 2] \\ &= k_3 \end{aligned}$$

Similarly, demand from the fourth ring is also k_3 . In short, although the total population in successive rings increases according to the square of the radius, a decay function of d^{-1} exactly offsets the expansion in area. Thus a value of $b=1$ provides a convenient basis for classification; if b is greater than 1 the absolute number of trips from successive rings will diminish outwards, if b is less than 1, it will increase outwards. As far as I am aware this useful property has not been noted in the literature on spatial interaction. In overlapping catchment areas which occur in reality, where preference as well as demand may decline away from a centre, it is very rare for trips to any one centre to increase outwards. In such circumstances a b value describing only a fall in preference without any fall in demand is therefore unlikely to be less than one. Within our restricted deterministic catchments, however, a b value of 1 or somewhat less would represent an appreciable drop in demand. It may be worth noting here that once catchments are allowed to overlap (i.e., the assignment of users to centres is probabilistic), it is no longer immediately clear whether a given b value represents elastic demand

or not. Under these circumstances to ascertain whether total demand from a relatively inaccessible area is less than that from an area close to a centre, we need to aggregate separately the trips from each area to all centres and compare them.

Returning to the general case for deterministic catchments, as long as b is not 2 ,

$$T_j = k_3 \int_0^a r^{1-b} .dr \quad (\text{from 3.5})$$

$$= k_3 \left[\frac{r^{2-b}}{2-b} \right]_0^a \quad (\text{from 3.6})$$

$$= \frac{k_3}{2-b} [a^{2-b}] \quad (3.8)$$

This expression gives the number of trips from one catchment of radius a . The number of trips from m catchments of radius a_m , denoted by $\sum_j T_j$, will be

$$\sum_j T_j = m. \frac{k_3}{2-b} [a_m]^{2-b} \quad (a_1)$$

$$= \frac{mk_3}{2-b} (a_1 m^{-1/2})^{2-b}$$

$$= \frac{mk_3}{2-b} (a_1^{2-b} m^{-1+b/2})$$

$$= \frac{k_3}{2-b} a_1^{2-b} m^{b/2}$$

$$= k_4 m^{b/2} \quad (3.9)$$

$$\text{where } k_4 = \frac{k_3}{2-b} a_1^{2-b}$$

Thus, by notionally unwrapping all m demand cones in the landscape, we have been able to express the relationship between m and aggregate demand in terms of the parameter b and a series of constants. Now, as b approaches 2 the exponent of m in 3.9 approaches 1 and the relationship between m and $\sum_j T_j$ becomes

almost linear. A linear relationship would mean that each successive facility stimulated the same amount of extra demand as the previous one. This surprising result arises apparently because, when b is nearly 2, use falls off so steeply from the centres that virtually all the demand comes from very close to each centre; each new facility is then set up in the middle of an area of almost untapped potential demand and therefore adds a linear increment to overall use (i.e. to $\sum_j T_j$). In short, demand is so elastic that each new centre practically generates as much extra demand as the first one itself. If one centre stimulates enough demand to support it comfortably, then it will likely be easy to justify many more centres.

The situation just described is unlikely to be true of the most essential public services, as they will probably have fairly inelastic demand. For them a saturation level will be reached fairly quickly with the increments in demand becoming progressively smaller as m increases. It is conceivable, however, that the growth in demand could be almost linear in the early stages of extending a new service or in a third world context where the cost and difficulty of travel could suppress demand very powerfully.

When b is exactly 2, we have to use 3.7 to integrate within one catchment, as noted earlier; in this instance 3.9 is no longer valid. Therefore within one catchment area with b equal to 2

$$\begin{aligned}
 T_j &= k_3 \int_0^a r^{1-b} .dr = k_3 \int_0^a r^{-1} .dr \\
 &= k_3 [\log_e r]_0^a && \text{(using 3.7)} \\
 &= k_3 [\log_e a - \log_e 0] = k_3 [\log_e a - (-\infty)]
 \end{aligned}$$

Thus the result is an expression which we can regard as infinite or undefined. When $b=3$ the integral is negative, which has no obvious interpretation in terms of demand. Retreating to firmer ground, we can

regard $b=2$ as a limiting case: as b approaches 2 the relationship is almost linear, but at 2 the whole relationship vanishes. Perhaps this helps to explain why the empirical studies mentioned above usually find a b value less than 2 .

Equation 3.9 defines a family of curves relating m and $\sum T_j$. The profile or shape of these curves depends only on b of course, not on the constants (a_1, k_2, k_3 and k_4). By selecting suitable arbitrary values for a, k_1 and k_2 we can define k_3 from 3.5.1 and thereby plot the curve for any given value of b less than 2 (Fig. 3.8). Since these curves outline the way demand responds to improvement in geographical access to supply, we can refer to them as demand response or simply user response curves.

The most striking feature of Figure 3.8 emerges from a comparison of the curves for the most elastic value of b , 1.9, and the least elastic, 0.5 . Naturally the less elastic curve has a higher level of demand when $m=1$ because it represents a situation where demand remains consistently higher throughout the whole catchment. However, when m is greater than two, the steep almost linear gradient of the very elastic curve quickly carries it well above the relatively inelastic one. Thus we have the result that demand in the system is higher when individual demand cones are steep and distance strongly suppresses demand from remoter locations, a profoundly illogical situation.

Further contradictions show up if we make a broader comparison of the curves. When $m=1$, the curve for $b=0.5$ is higher than that for $b=1.0$ which in turn is higher than that for $b=1.5$, as we would expect. Instead of being lower again than the latter, however, the curve for $b=1.9$ is actually slightly higher than the latter two.

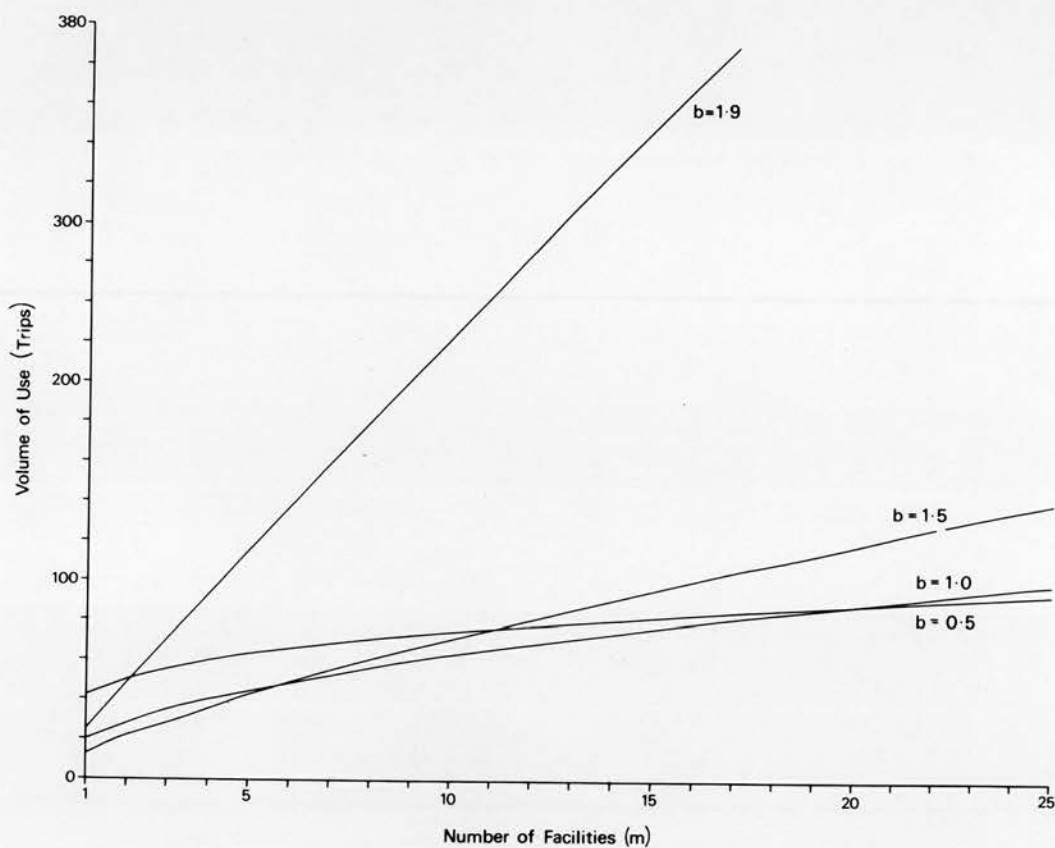


Figure 3.8 Relation between volume of use and m on an isotropic plane for various values of b (user response curve)

For higher values of m the relative positions of the curves for $b=0.5$, $b=1.0$ and $b=1.5$ are eventually inverted with the most elastic curves reaching higher levels beyond $m=20$.

Because these results seemed somewhat illogical, the program used to compute them, GENEVA3, was checked very carefully; selected values were also computed by hand to confirm their validity. In fact, as we will argue below, these paradoxical results are an intrinsic property of this particular model of elastic demand. The remedy is therefore to seek a model with fewer contradictions.

More flexible types of demand cone

The difficulties encountered in trying to define a user response curve for $b=2$ highlight a general drawback of this type of formulation: even with low values of b a function of the type $1/b^d$ climbs very steeply as it approaches the y-axis because it is inherently asymptotic to the latter axis (Fig.3.9) . Taylor (1975) notes that curves of this type tend to overpredict the amount of spatial interaction at short distances, when fitted to observed data by regression methods. He explains this by saying that the logarithmic transformation employed for regression "overtransforms" the data past linearity so that the curve is concave upwards. However it could also be argued that this discrepancy stems not from the transformation but from the asymptotic nature of the function itself. This problem can be solved by using a function of the form $1/(1+d^b)$ or by using a negative exponential function such as e^{-bd} . Both intersect the y-axis when $d=0$ and therefore avoid the problems associated with being asymptotic. Beyond short distances, however, the former function is practically the same as $1/b^d$ and, being asymptotic to the x-axis, would probably tend to overpredict at longer distances, as appears to happen in the examples presented by

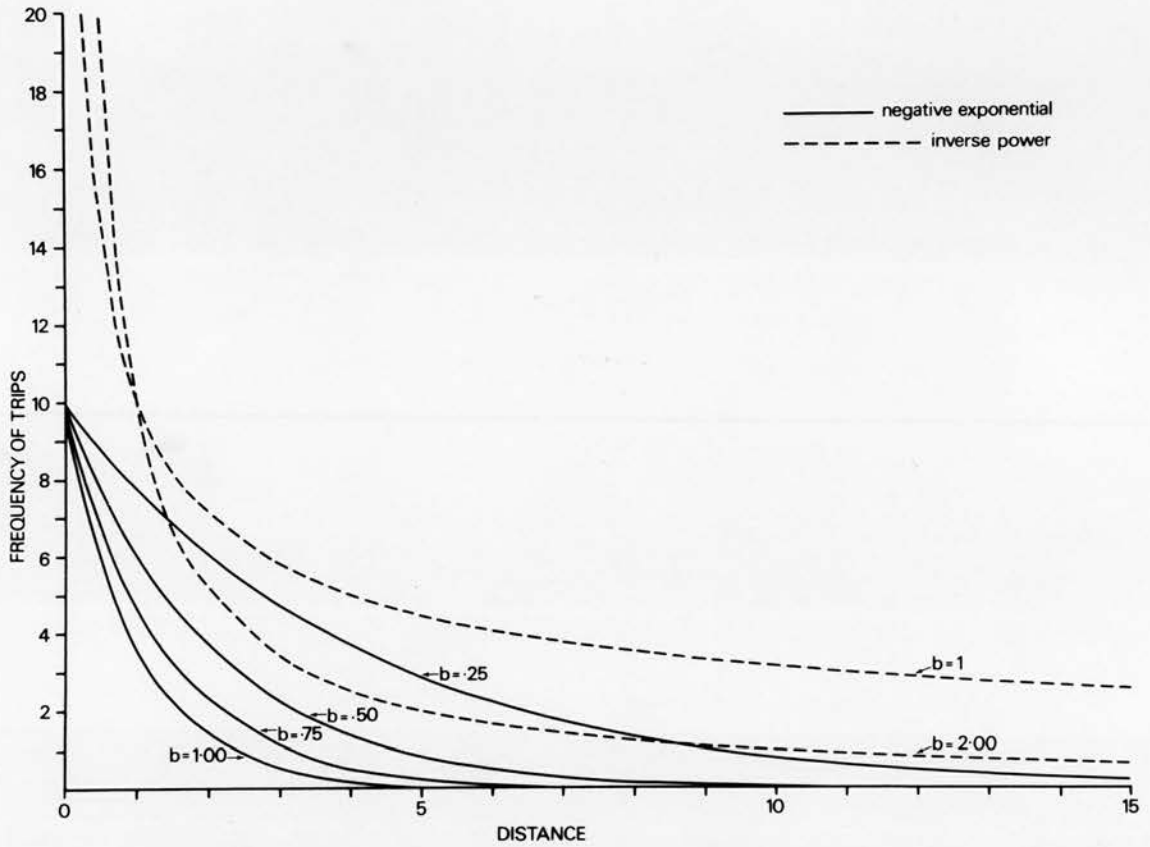


Figure 3.9 Comparison of distance decay curves based on inverse powers and on negative exponentials

Taylor. In contrast a function of the type e^{-bd} is practically zero when b and bd are very large. These properties help to explain, perhaps, why negative exponentials give better fits to the data in Taylor's examples.

The contradictions encountered earlier arose therefore mainly because the function employed was asymptotic to the y-axis, implying a high volume of demand under the curve near the origin. In fact the larger the value of b , the greater the demand from distances less than 1 km. This explains why some of the demand response curves have their relative positions inverted when m is so large that catchments are small enough to be dominated by demand from distances less than one km.

If f declines in a negative exponential manner and f_0 is the frequency of trips from an area so close the centre that $d=0$ then

$$f = f_0 \cdot e^{-bd^n} \quad (3.10.1)$$

where n is a parameter of the relation, like b . For convenience and easier comparison we can take $f_0 = k_2$ as defined earlier in (3.4.2). We then have

$$f = k_2 e^{-bd^n} \quad (3.10.2)$$

This formula defines a family of demand cones each specified by particular values of b and n . The greater the values of b and n the steeper the cone. In Taylor's examples values of $n=1$ and $n=\frac{1}{2}$ yielded good fits at short and long distances and had the smallest standard error of the estimate. Since $n=1$ is simpler, it will be used here in preference to $n=\frac{1}{2}$. In terms of the variety of demand cones we are able to treat, little is sacrificed by ignoring the latter value. We can now compare a family of negative exponential curves with some of the functions used earlier; the advantages of the

former at long and short distances are readily apparent (Fig.3.9) .

We can now try to show how use will respond to an increase in m when the demand cones have a negative exponential form.

The number of trips from an annulus, i , r units distant from the centre of a circular catchment will now be the product of f and the ring's population. Hence

$$T_i = k_2 e^{-br} \quad 2\pi r \cdot dr \cdot p. \quad (3.11.1)$$

The number of trips from the whole catchment will therefore be

$$\begin{aligned} T_j &= k_2 2\pi p \int_0^a r e^{-br} \cdot dr \\ &= k_3 \int_0^a r e^{-br} \cdot dr \end{aligned} \quad (3.11.2)$$

There is no standard integral for an expression of the type $x e^{-bx}$. However, we can use the method of integration by parts which exploits the following relation:

$$\int u \frac{dv}{dx} \cdot dx = uv - \int v \frac{du}{dx} \cdot dx \quad (3.12)$$

where $f(x) = u(x) \cdot v(x)$

If we define $u = x$ and $v = e^{-bx}$

then $\frac{du}{dx} = 1$; $\frac{dv}{dx} = e^{-bx} (-b)$; and of course $uv = x e^{-bx}$

Substituting in 3.12

$$\begin{aligned} \int x(-b)e^{-bx} \cdot dx &= x e^{-bx} - \int e^{-bx} \cdot 1 \cdot dx \\ \therefore -b \int x e^{-bx} \cdot dx &= x e^{-bx} - (-1/b e^{-bx} + c) \\ \therefore -b \int x e^{-bx} \cdot dx &= e^{-bx} (x + 1/b) - c \\ \therefore \int x e^{-bx} &= -1/b e^{-bx} (x + 1/b) - c . \end{aligned}$$

By replacing x with r we can now integrate 3.11.2 .

$$\begin{aligned} T_j &= k_3 \left[-1/b e^{-br} (r + 1/b) - c \right]_0^a \\ &= k_3 \left[-1/b e^{-ba} (a + 1/b) + 1/b \cdot 1/b \right] \\ &= k_3 b^{-1} [-e^{-ba} (a + b^{-1}) + b^{-1}] \end{aligned} \quad (3.13)$$

Equation 3.13 gives us an expression for the demand from one catchment area of radius a . The total demand from all m catchments, each of radius $a_m (= a_1 m^{-1/2})$, is therefore

$$\sum T_j = m k_3 b^{-1} [-e^{-ba_m} (a_m + 1/b) + 1/b] \quad (3.14)$$

This expression defines a second family of demand response curves. As before we can evaluate the expression for a range of demand cones by taking any suitable value of k . Because the form of our function is now different, we need to take a somewhat different range of b values from before. We have found it convenient to use $b = 0.25$ to define a relatively inelastic cone and $b = 1.00$ to define one which is relatively elastic with intermediate values of $b = 0.5$ and $b = 0.75$ (Fig.3.10).

In absolute terms demand is always greater when elasticity is low (i.e., the cone has a gentle slope) but tends to level off more quickly (Fig.3.10). Conversely when elasticity is high, the curve is lower but more linear. Hence the increments to demand diminish less rapidly in the latter instance. It is also worth noting that these demand response curves are free of the contradictions encountered with the inverse power model.

Equation 3.14 and Figure 3.10 could form the basis of a method for projecting how the demand for a service might rise as access to supply improved - a question which seems to have particular relevance to indoor recreation. If we had data on the way in which distance or cost inhibited demand to a relatively limited number of existing centres, we could estimate a b value for that data. By entering this b value in a more detailed version of Figure 3.10, it would then be possible to gauge very roughly how much "latent" demand existed for the particular service with the number of centres in existence.

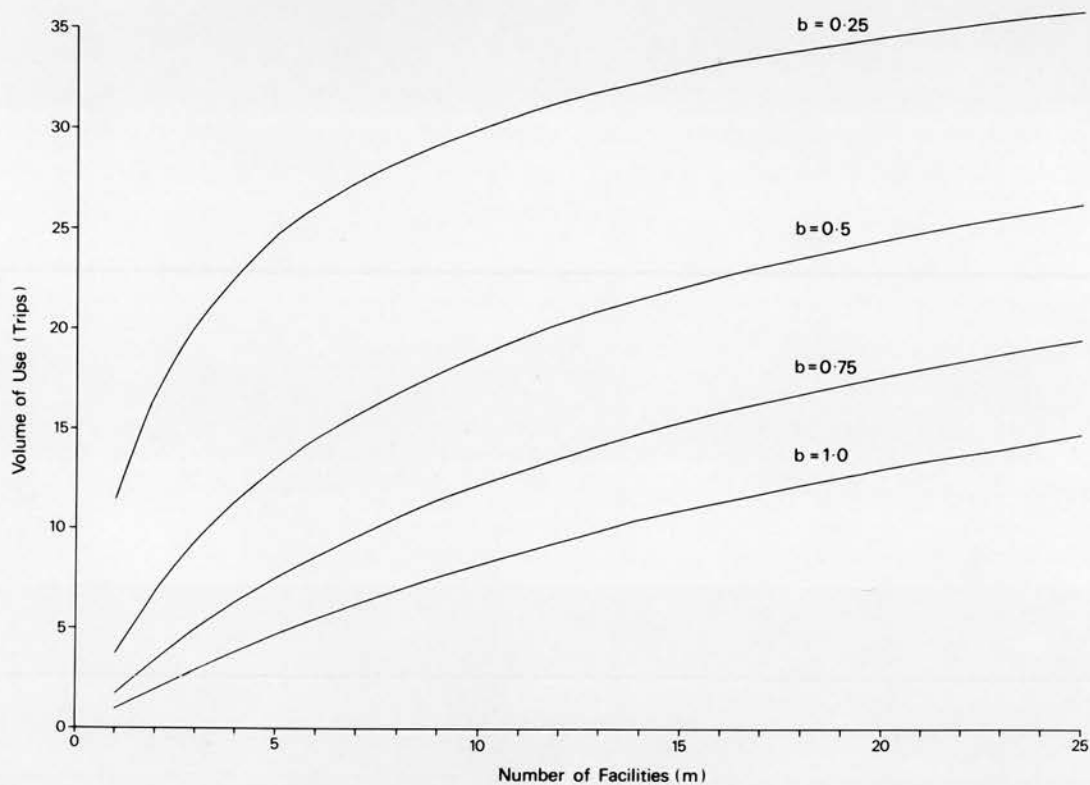


Figure 3.10

Family of user response curves with decline in use
described by a negative exponential function -
absolute values

In making such a projection the underlying assumptions would need to be kept in mind. In particular, two assumptions should be noted. Firstly, the more a given transport network departs from the pattern of travel costs imposed by a uniform plane, the less valid would this projection be. Under these circumstances, a b value could still be estimated but a more heuristic, less analytical procedure than the present one would be needed to reveal the implications for growth in demand. It should, however, be relatively easy to translate such a method into a computer algorithm. A second problem arises from ignoring the time dimension and from the use, initially, of a Löschian landscape, instead of Christaller's one. Briefly, when a fairly large number of facilities has been built, in the real world it will often be better to locate any extra ones in their interstices, which violates the assumptions underlying our model. This raises the problem of where to locate additional centres in a real landscape - the problem of sequencing, which will be mentioned in a later chapter.

Though we largely ignore the question of interstitial locations in the present chapter, it might be possible to throw some light on this problem by adapting the methods developed here. Suppose we ask how much demand would be generated by an extra facility located in the interstice between m uniformly spaced existing centres. If demand is so elastic that most users come from short distances, the new facility will generate roughly the same amount of demand as the existing centres already attract. Similarly if demand is fairly inelastic but we allow demand cones to have extensive overlaps, as we will do in the next section, the same is more or less true, though all the centres will have somewhat less demand after the new one is introduced.

When catchments are deterministic the demand attracted by the new facility itself is easy to define. From Christaller's $k = 3$ landscape

we know the area of the new catchment will be one third the area of the other catchments; it will therefore have a radius $1/\sqrt{3}$ times their radii i.e. $a_m/\sqrt{3}$. Its demand can thus be obtained by integrating 3.11.2 between the limits of zero and $a_m/\sqrt{3}$. The three facilities adjacent to the new centre, however, will have catchments truncated to a radius of $a_m/\sqrt{3}$ on that side but still extending to a_m on the other sides. Thus we have the problem of estimating demand under a non-symmetrical cone, which appears much more difficult than the problems already tackled. A method could probably be devised to do so, but its derivation is outside the scope of the present work.

Despite the practical limitations just listed, equation 3.14 does provide an answer to the basic theoretical question about the relationship between the number of centres and the level of use, an answer which is more satisfactory logically than that provided by equation 3.9. In a sense it can therefore be seen as an extension of Lösch's central place theory and as an attempt to make certain aspects of central place theory more explicit mathematically. On a more practical level, Figure 3.10 could be quite useful as a kind of reference chart to be employed in exploring the implications for the whole system of any given rate of distance decline.

The user response curves can be compared in relative as well as absolute terms by standardising their values so that the volume of demand with one centre in existence is scaled to 100 (Fig.3.11). Standardising in this way can be misleading as it inverts the magnitudes in absolute terms but it does make comparison of slopes easier and therefore emphasizes more strongly the difference in gradient and the almost linear character of the more elastic response curves. The latter curves afford a striking contrast to the steadily

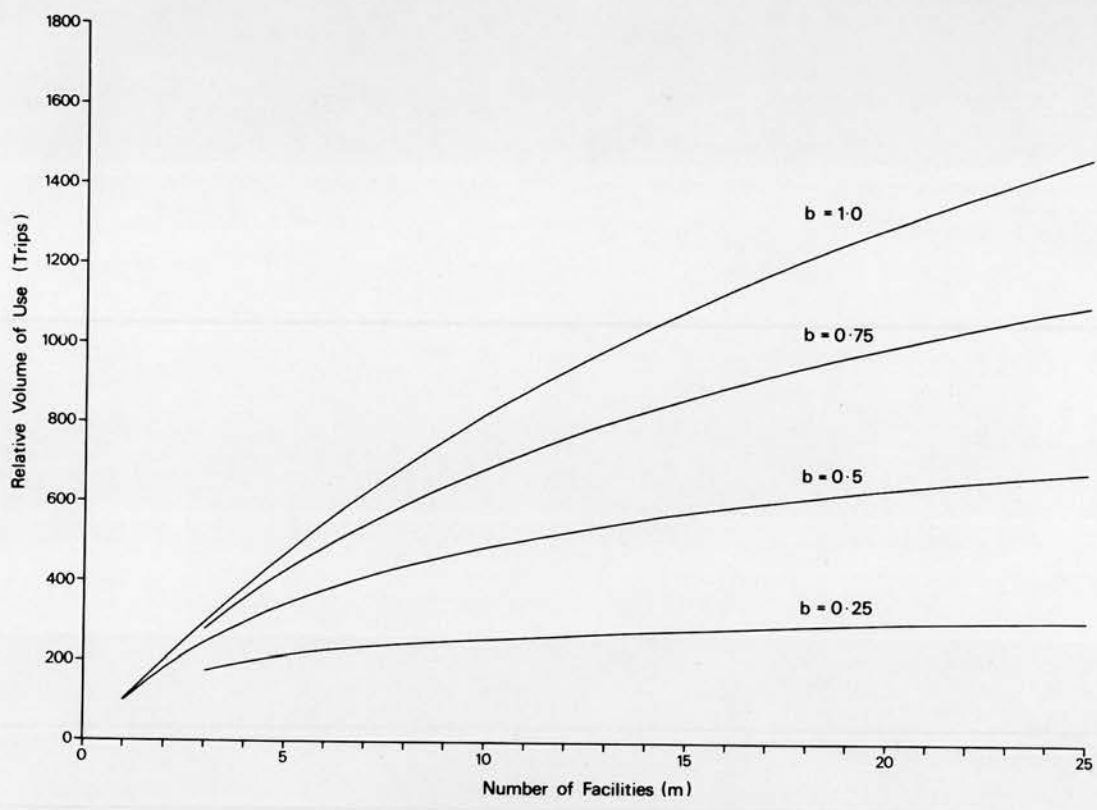


Figure 3.11

Family of user response curves with decline in use described by a negative exponential - relative to values with $m \neq 1$

diminishing marginal benefit to users for the AT and MT curves with inelastic demand, noted earlier (Fig.3.4) . The relative curves would also be more convenient to use in an exercise concerned with projecting demand: given demand for one centre and a genuine b value for that centre, we can refer to the graph and thereby project demand in percentage terms for a given value of m .

Hitherto we have ignored the question of whether a_1 , the shorter radius of the largest catchment, has any influence on our results. Consider now an area of radius 10 km ($a_1=10$) . This area would be slightly larger than the area of the City of Edinburgh before local government reorganisation. Suppose the demand cone for a certain service is defined by $b=0.5$; the nature of this curve is such that the number of trips from beyond 8 km is negligible. Obviously the number of cones with restricted catchments which can be fitted into this space before successive increments fall off sharply is relatively small. However, within an area of radius 25 km (slightly larger in total surface than Lothian Region) obviously a greater number of cones could be accommodated before the absolute increments fell below a specified level. For this reason a_1 could be regarded as a parameter of the relationship. Despite its influence, it has not been necessary to take a_1 directly into account. Substituting $a_1 m^{-\frac{1}{2}}$ for a_m we can rewrite (3.14) as follows:

$$T_j = m k_3 b^{-1} [-e^{-ba_1 m^{-\frac{1}{2}}} (a_1 m^{-\frac{1}{2}} + b^{-1}) + b^{-1}] \quad (3.15)$$

It seems reasonable to suppose that the shape of the curves in Figure 3.11 will depend on the size of the demand cones, defined by b , relative to the total area, defined by a_1 ; hence the profiles might depend roughly on the product ba_1 . Now the most influential term in equation 3.15 is $-e^{-ba_1 m^{-\frac{1}{2}}}$ in that it responds most

dramatically to changes in b or a_1 ; the fact that this term is controlled by the product ba_1 lends further weight to this broad supposition. Experiments with various values of b and a_1 showed that the user response curve expressed in relative terms appears to depend precisely on ba_1 as Table 3.1 shows for a particular case. The author cannot prove by mathematical means why this should be exactly so. Whatever its proof, this result made it unnecessary to explore the influence of a_1 separately. Thus in computing all the results presented earlier a_1 was given a standard value of 10 km, while b took on the standard range of values.

Table 3.1 Demand response values expressed in relative terms for different values of a_1

$a_1 = 10$; $b = 1.0$

$ba_1 = 10$

<u>m</u>	<u>Demand Response</u>
1	100.0
2	198.7
3	293.8
4	384.0
5	468.9
6	548.8
7	623.9
8	694.6
9	761.2
10	824.2
11	883.8
12	940.4
13	994.1
14	1045.3
15	1094.2
16	1140.8
17	1185.5
18	1228.3
19	1269.4
20	1308.9

$a_1 = 5$; $b = 2.0$

$ba_1 = 10$

<u>m</u>	<u>Demand Response</u>
1	100.0
2	198.7
3	293.8
4	384.0
5	468.9
6	548.8
7	623.9
8	694.6
9	761.2
10	824.2
11	883.8
12	940.4
13	994.1
14	1045.3
15	1094.2
16	1140.8
17	1185.5
18	1228.3
19	1269.4
20	1308.9

All of the preceding argument on elastic demand depends on using inscribed circles to approximate hexagonal catchments. Even when demand falls only moderately, proportionately less demand comes from

the relatively distant part of the hexagon outside the inscribed circle than with fixed demand. Hence the approximation is more accurate here than in the model with inelastic demand; indeed the estimate will be very accurate when demand is strongly elastic.

Elastic demand with overlapping catchment areas

As in classical central place theory, the catchment areas treated so far have been deterministic in that they have not been allowed to overlap, in an unsatisfactory limitation. It is fairly easy to relax this restriction mathematically. Hitherto, in each catchment we have found the amount of demand between the centre and the boundary of the hexagon by integrating between the limits zero and a ($=a_m$). However, we could find the amount of demand which lies beyond the boundary of the hexagon by integrating to an upper limit greater than a , $2a$ or $3a$ for instance. Our catchments then overlap; the assignment of users to facilities is now on a probabilistic basis.

Only a small change is needed in the computer program written to evaluate 3.14 to put this into effect. After demand has been computed up to a limit of a , we simply redefine a as $2a$, $3a$ etc. and repeat the evaluation. In this context it is convenient to define an 'overlap factor' by taking the upper limit used in integration as a ratio of a_m : an overlap factor of two means that the upper limit is $2a_m$. As might be expected, this modification makes relatively little difference to the results for steep cones since so little of their demand lies beyond a_m , as long as a_m is not very small (Fig.3.12). On the other hand it makes an appreciable difference to the cones with gentle slopes (Fig. 3.13), because their catchments can potentially come from a much wider area. The user response curves derived in this way can therefore be regarded as a more satisfactory basis for projecting moderately elastic demand and thus represent a useful extension of

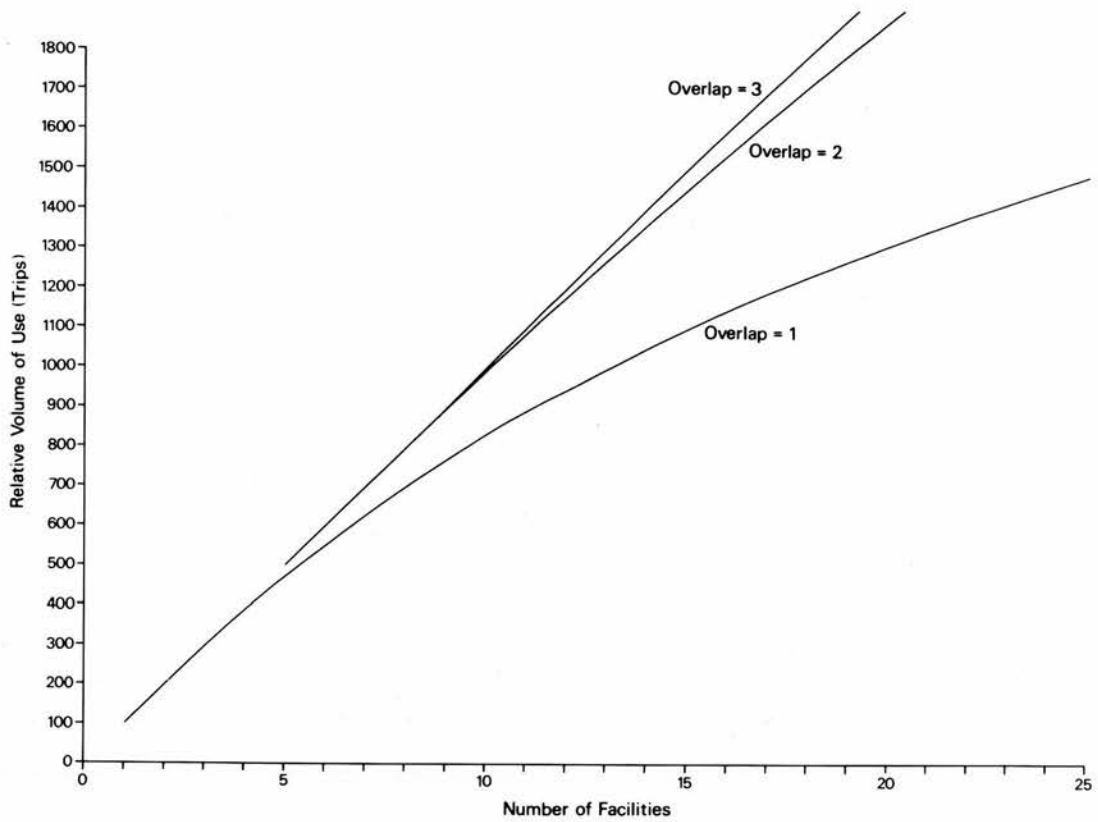


Figure 3.12

Effect of overlapping catchments on user response curves when demand is strongly elastic ($b = 1$)

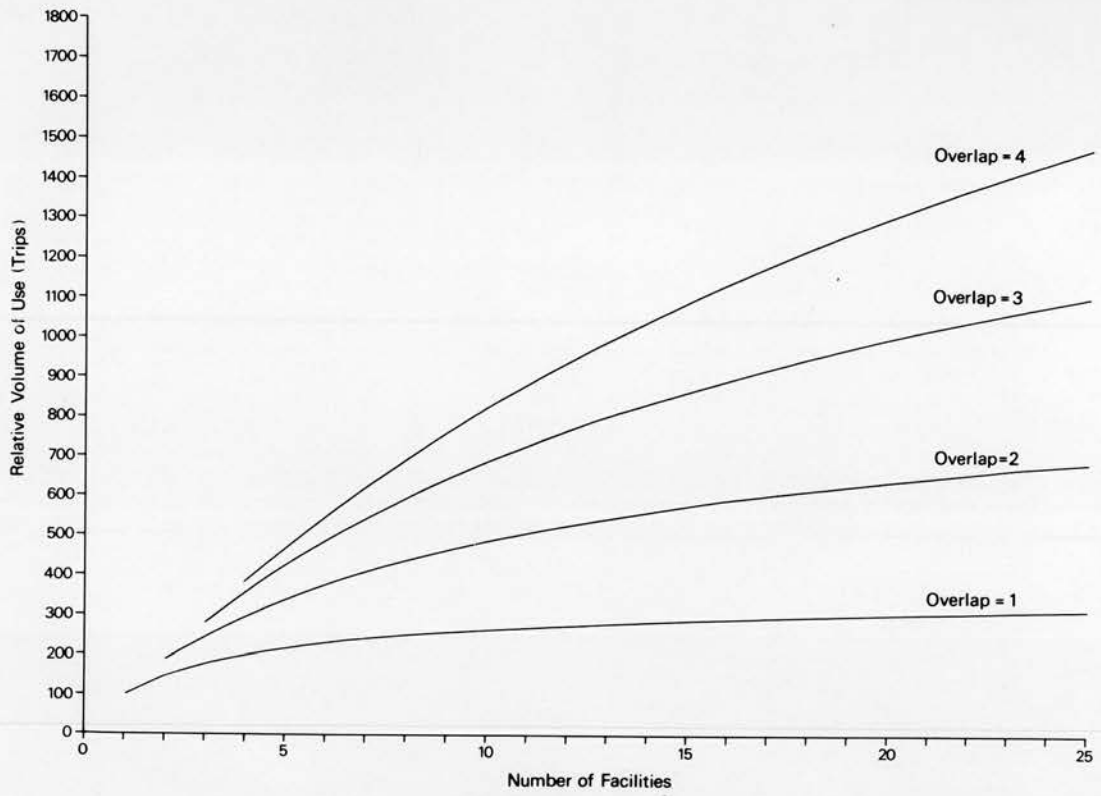


Figure 3.13 Effect of overlapping catchments on user response
curves when demand is mildly elastic ($b = .25$)

the model.

As noted earlier, the profile of the user response curve expressed in relative terms depends on the product $b \cdot a_m$. An overlap factor of 2 therefore implies doubling this product which has the same relative effect as halving the slope b . Thus the curve produced by $b=0.5$ and overlap = 1 has the same relative shape as that for $b=0.25$ and overlap = 2. In general, curves with the same product of overlap and gradient must have the same profiles when these are expressed in relative terms.

As noted earlier, when catchments overlap it is no longer immediately clear whether aggregate demand declines with distance or not. In fact it may even be conceivable that by specifying a gentle cone and allowing extensive overlap, we could generate more demand at a point in the interstices between centres than at a point adjacent to one centre. There are several problems concerned with the general level of demand lurking in the background here which lie outside the scope of the present work. Nevertheless, it would always be possible to test whether this problem had arisen by writing a program to check if the level of demand in the middle of an interstice exceeded that at a facility itself.

Although a satisfactory analytical method has not yet been devised, there is a rough means of testing for this possibility. When demand falls as e^{-bd} , the number of trips at $d=0$ is f_0 or k_2 . If demand was inelastic this would be the frequency of trips from the whole area. Since the whole area's population is $2\sqrt{3} a_1^2 p$, the total number of trips will be $f_0 2\sqrt{3} a_1^2 p$. When catchments overlap we can test whether this number has been exceeded at any point as more centres are provided. If so the assumption of elastic demand has in

a sense been violated. Thus, with deterministic catchments we can take $f_0 \cdot 2\sqrt{3} a_1^2$ as a measure of the potential demand available and compute the actual demand released at any point as a ratio of it.

As Table 3.2(a) shows, when demand is highly elastic ($b=1.0$) only 36.3% of the potential demand available is released by a supply of twenty facilities, even though the overlap factor is 4.0. On the other hand with $b=0.25$ the theoretical maximum level of demand is exceeded with an overlap of only two and a supply of six facilities (Table 3.2(b)), so our underlying construct of elastic demand has apparently been violated. There is little point in examining this problem in greater depth here but if we allow catchments to overlap, we must be aware of this difficulty.

With overlapping catchments it becomes more difficult to ignore the boundary problem occurring at the edge of the whole landscape. For instance, given one centre, an overlap of 2 and a gentle demand cone, we are in effect importing demand from beyond the boundaries of the landscape; demand is exported there as well. Though the relative amounts imported and exported will diminish as m increases and catchments become smaller in size, the results for low values of m under such circumstances have to be viewed rather critically. When catchments are discrete, this boundary problem can be regarded as almost negligible, though it does exist.

In conclusion, probabilistic catchments introduce some extra difficulties, though, essentially, they make the present framework more flexible and comprehensive. The problems exposed here could form the basis for further work.

Elastic demand has been treated at some length partly because it is a more complex topic and partly because it has been neglected in previous work. In fact because we can now express the impact of more

centres on demand and aggregate travel cost, it is possible to synthesise both within a more general framework, as we will argue later.

Table 3.2 Demand Response Values Expressed as a Percentage of Total

Demand Available

(a) Strongly Elastic Demand : $b = 1.00$

<u>Overlap = 1.00</u>			<u>Overlap = 4.00</u>		
<u>m</u>	<u>Demand</u>	<u>Percentage of Total</u>	<u>m</u>	<u>Demand</u>	<u>Percentage of Total</u>
1	1.000	1.813	1	1.000	1.814
2	1.986	3.603	2	2.000	3.628
3	2.937	5.327	3	3.000	5.441
4	3.838	6.962	4	4.000	7.255
5	4.687	8.502	5	5.000	9.069
6	5.486	9.950	6	6.000	10.883
7	6.236	11.311	7	7.000	12.697
8	6.943	12.592	8	8.000	14.510
9	7.609	13.801	9	9.000	16.324
10	8.238	14.942	10	10.000	18.137
11	8.834	16.023	11	10.999	19.950
12	9.399	17.049	12	11.999	21.763
13	9.937	18.023	13	12.998	23.575
14	10.449	18.952	14	13.996	25.386
15	10.937	19.837	15	14.994	27.197
16	11.403	20.683	16	15.992	29.006
17	11.850	21.493	17	16.989	30.814
18	12.278	22.269	18	17.985	32.621
19	12.688	23.014	19	18.980	34.426
20	13.083	23.729	20	19.974	36.229

Table 3.2(b) Mildly Elastic Demand : $b = 0.25$ Overlap = 1.00

<u>m</u>	<u>Demand</u>	<u>Percentage of Total</u>
1	11.403	20.683
2	16.880	30.617
3	20.306	36.831
4	22.743	41.252
5	24.605	44.629
6	26.095	47.332
7	27.326	49.565
8	28.368	51.454
9	29.266	53.083
10	30.052	54.508
11	30.747	55.769
12	31.369	56.896
13	31.929	57.913
14	32.438	58.837
15	32.904	59.681
16	33.331	60.456
17	33.726	61.173
18	34.092	61.837
19	34.433	62.455
20	34.752	63.033

Overlap = 2.00

<u>m</u>	<u>Demand</u>	<u>Percentage of Total</u>
1	15.353	27.848
2	27.770	50.370
3	37.598	68.195
4	45.613	82.733
5	52.331	94.918
6	58.084	105.353
7	63.095	114.442
8	67.521	122.469
9	71.472	129.636
10	75.033	136.094

Net benefit to users

Within the present framework the net benefit to a user of one trip to a centre can be defined as the benefit or utility he enjoys from using the service (a game of squash or a visit to a library, for instance) minus the travel cost of his journey. Hence for one trip by one user:

$$\text{net benefit} = \text{benefit or value of service} - \text{travel cost.}$$

Consider a situation where demand is elastic within a system of Thiessen catchments. If the benefit in monetary terms to each individual of each trip is v units, the value to the whole population of $\sum_j T_j$ trips is therefore $v \sum_j T_j$, which we can denote by BU . The total travel cost for all users, TC will be the number of journeys multiplied by mean length of trip and the travel cost per km or, more simply, aggregate travel times travel cost per km. The net benefit to all users, NBU , will then be the difference between benefits and costs over the whole system. Thus

$$\begin{aligned} NBU &= BU - TC \\ &= v \sum_j T_j - \sum_j T_j \cdot MT \cdot c_t \end{aligned} \quad (3.16)$$

where MT is mean length of trip for all users

c_t is the cost of travel per km.

Alternatively, replacing $\sum_j T_j \cdot MT$ by AT

$$NBU = v \sum_j T_j - AT \cdot c_t$$

It is convenient to note at this point that BU is the product of a constant, v , and a variable, $\sum_j T_j$. We will call the latter **TRIPS** for convenience.

To evaluate NBU , we need to know either AT or MT . Our previous derivation of AT and MT applied only to inelastic demand. We now need to derive an expression for AT and MT when demand is

elastic, a rather more difficult task. When demand declines as e^{-bd} , it proves easier to derive AT.

Consider again a single catchment and an annulus, i , within it. Total travel from i will be the number of trips, given by 3.11.1, times the distance to the centre, r ; hence

$$AT_i = k_2 e^{-br} \cdot 2\pi r \cdot dr \cdot p \cdot r$$

Aggregate travel from the whole catchment can then be found by integration:

$$AT_j = k_2 2\pi p \int_0^a r^2 e^{-br} \cdot dr$$

We can integrate $r^2 e^{-br}$ with respect to r by using the method of integration by parts in a manner similar to that used for $r e^{-br}$ (see 3.12). This yields the result

$$\begin{aligned} AT_j &= k_2 2\pi p [-b^{-1} e^{-br} (r^2 + 2rb^{-1} + 2b^{-2}) + 2b^{-1}c]_0^a \\ &= k_3 [-b^{-1} e^{-ba} (a^2 + 2ab^{-1} + 2b^{-2}) + b^{-1} \cdot 1 \cdot (2b^{-2})] \end{aligned}$$

Hence AT for m catchments of radius a_m is

$$\sum_j AT_j = m k_3 [-b^{-1} e^{-b a_m} (a_m^2 + 2a_m b^{-1} + 2b^{-2}) + 2b^{-3}]$$

Mean travel over the whole system will be the same as within each catchment. As before it can be found by dividing AT by the number of trips.

It is difficult to decide what value should be given to v . Estimating v could be a research project in itself or it could even be regarded as fairly intractable philosophical problem. Any value will be arbitrary to some extent; furthermore it can be argued with some reason that it is impossible to assign a monetary value to the benefit derived from visiting a doctor or a concert hall. Nevertheless, we can argue that the real value of any visit must at least exceed the travel cost or the journey would not be made. Since our main aim here is to

understand the nature of the net benefit curve we can explore its profile by assigning a range of values to v from high to low.

Let us suppose that at least one individual travels the maximum distance of a_1 units to use the service when there is only one centre. The value of one trip to him must then at least exceed $a_1 c_t$, the cost of his journey. We can therefore take $a_1 c_t$ as a relatively high value for v . A relatively low value could be obtained from $a_m c_t$, the cost of the longest journey when m facilities are in existence. The disadvantage of the latter, however, is that it declines as more facilities are added; which is logically unsatisfactory because it seems more plausible to assume that v will be constant with respect to m . It may therefore be preferable to simply assign an arbitrary low value to v . Since a_1 has been given a standard value of 10, if we take $c_t = 1$ then $v = 5$ may serve as a comparatively low value.

Figures 3.14 to 3.17 show the profiles of the indices, NBU, AT, MT and BU for the four standard values of b with $v = 10$. Each index is expressed relative to its own value when $m=1$, which is scaled to 100 to facilitate comparison. In all cases NBU rises more steeply than BU because BU measures only the increase in use; it does not take account of the concomitant reduction in travel cost. For the same reason NBU will rise more steeply than BU in absolute as well as relative terms. Thus we now have a framework which takes account of the fact that the increase in demand alone underestimates the true gain from adding more facilities.

The relative profiles of the aggregate travel curves are particularly interesting. AT is not, incidentally, dependent on the arbitrary value of v . When demand is very elastic ($b=1$) within

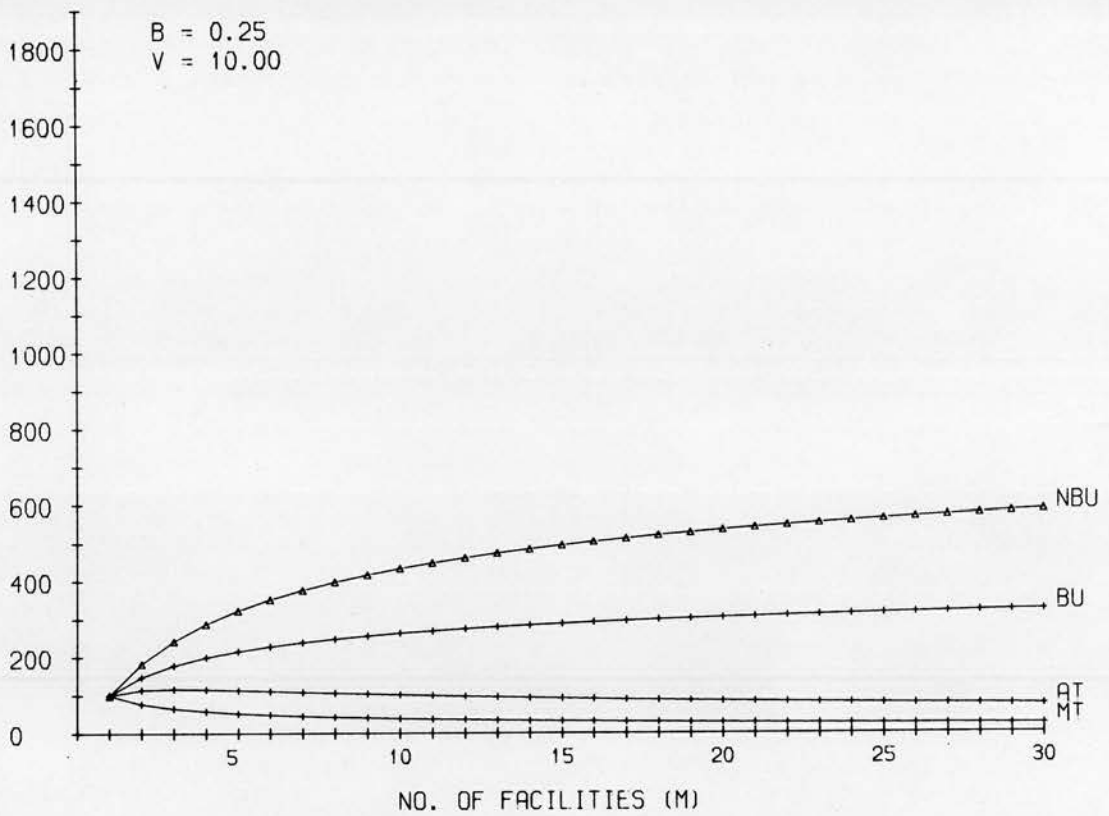


Figure 3.14

Relation between m and NBU, BU, AT and MT, expressed as percentages of values for $m = 1$, on an isotropic plane with $b = .25$ and $v = 10$

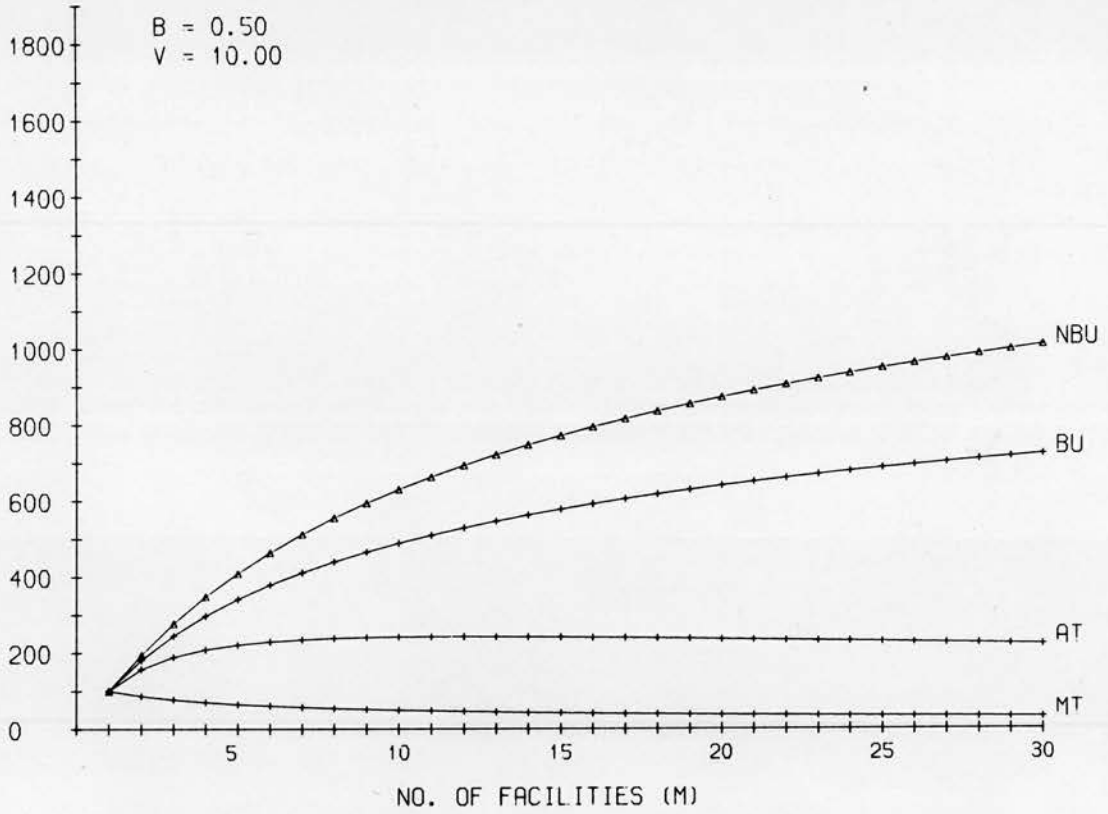


Figure 3.15

Relation between m and NBU, BU, AT and MT, expressed as percentages of values for $m = 1$, on an isotropic plane with $b = .50$ and $v = 10$

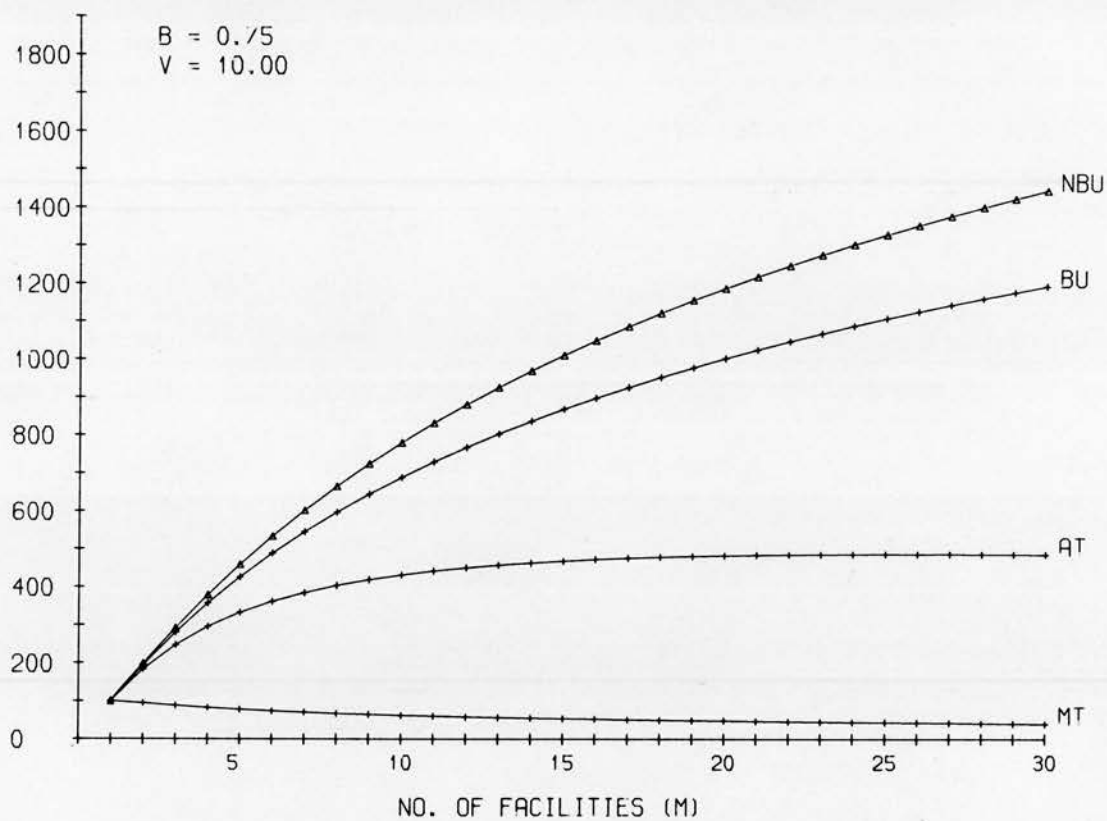


Figure 3.16

Relation between m and NBU, BU, AT and MT, expressed as percentages of values for $m = 1$, on an isotropic plane with $b = .75$ and $v = 10$

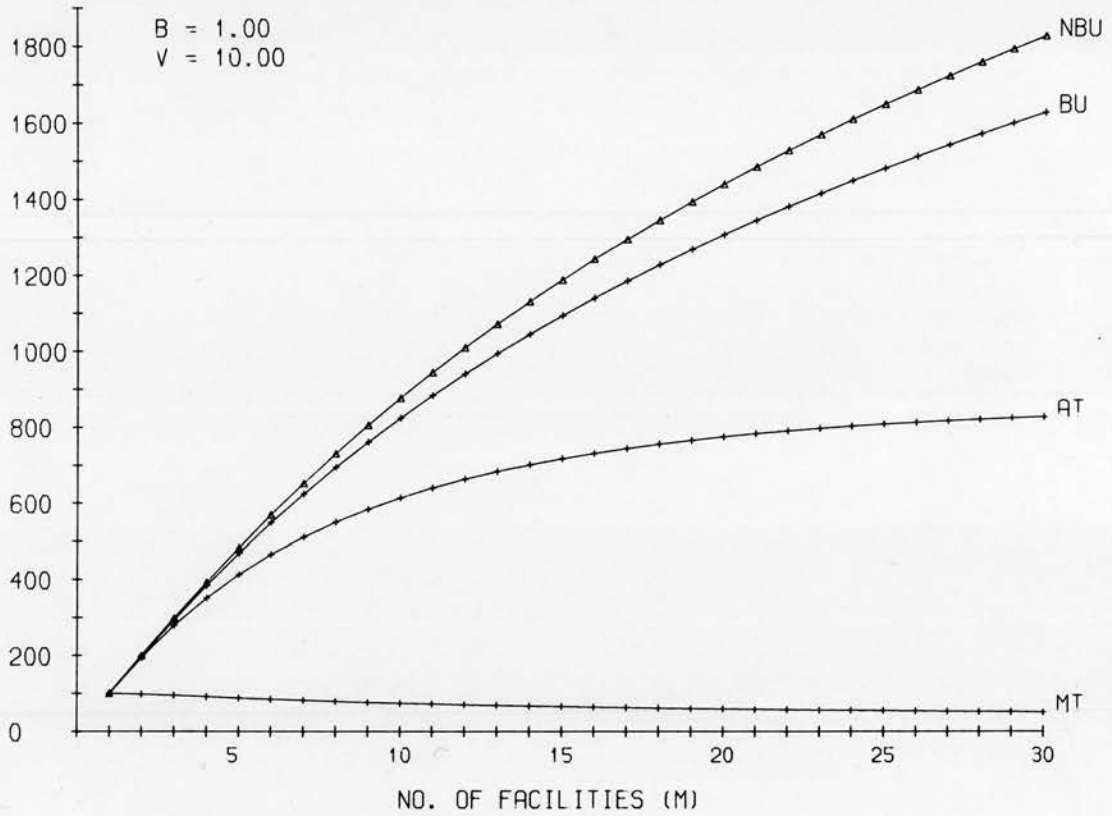


Figure 3.17

Relation between m and NBU, BU, AT and MT, expressed as percentages of values for $m = 1$, on an isotropic plane with $b = 1.0$ and $v = 10$

our range of m , AT rises continuously, simply because many more people are travelling to use the extra facilities. Conversely, when demand is relatively inelastic ($b=0.25$), aggregate travel falls at a steady, gentle rate after a small initial rise; thus with little extra demand being generated, the gains to users come mainly in the form of reduced travel costs. The value of b for which the AT curve lies fairly flat is probably just slightly above $.25$, which can be regarded as a kind of critical value: when b is greater most of the benefit comes from increased demand; below 0.25 it comes mostly from reduced travel cost. Indeed, we can now see the AT curve already computed for inelastic demand as a special case defined by $b=0$; NBU then consists only of travel savings and the profiles of AT and MT are identical. In contrast, a highly elastic demand produces the opposite limiting case in which virtually all the benefit is in the form of increased use. The third index, mean travel (MT), is a helpful measure of the distance travelled by users. In all cases it falls continuously but the gradient is steepest of course when the demand curve has a gentle slope.

Before examining the corresponding results for $v = 5$, we have to resolve a logical difficulty encountered when v is less than 10 if we continue to use equation 3.14 to define demand. To obtain 3.14 we integrated 3.11.2 between zero and an upper limit of a .

So far we have always taken this upper limit, a , to be a_m , the length of the longest trip in any catchment. The cost of such a trip is therefore $a_m c_t$ which is equal to $a_1 m^{-\frac{1}{2}} c_t$ (from 3.1). If our value of v is less than $a_1 m^{-\frac{1}{2}} c_t$, we are making the unsatisfactory assumption that a trip will be made when its travel cost exceeds its value.

To prevent this we should eliminate trips from beyond a distance

of a' where a' is such that

$$v = a'c_t$$

$$\text{i.e. } a' = v/c_t$$

Here, since $c_t = 1$ we must somehow prevent trips beyond a radius of $a' = v/c_t = v$. If we take $v = 5$, the longest permissible journey is thus 5 km. Now we can enforce this constraint by reducing the upper limit for integrating 3.11.2 from a_m to 5 km, when a_m exceeds 5. More generally, we can make the upper limit v/c_t . Hence the program written to evaluate and plot NBU (NETPL27) was modified to check whether v exceeded $a_m c_t$ at any stage. If it did, then the upper limit for integration, was changed from a_m to v/c_t . This modification also ensured that total travel distance $\sum_j AT_j$, was counted only for worthwhile journeys. With the standard values for constants this whole difficulty does not arise when v is greater than 10 because v then exceeds the maximum travel cost which is $a_1 c_t = 10$.

For a given value of v less than 10 this problem only arises of course when catchments are relatively large. It therefore tends to diminish in all cases as m increases, disappearing exactly when m is such that

$$a_1 m^{-\frac{1}{2}} c_t = v$$

$$\text{i.e. } m = \left(\frac{v}{a_1 c_t} \right)^{-2}$$

When $v = 5$, this expression yields $m = 4$. Only values of m below this critical level will be affected by the alteration in the program.

Figures 3.18 to 3.21 show the profiles of NBU, AT, MT and BU, again expressed in relative terms, for the four standard

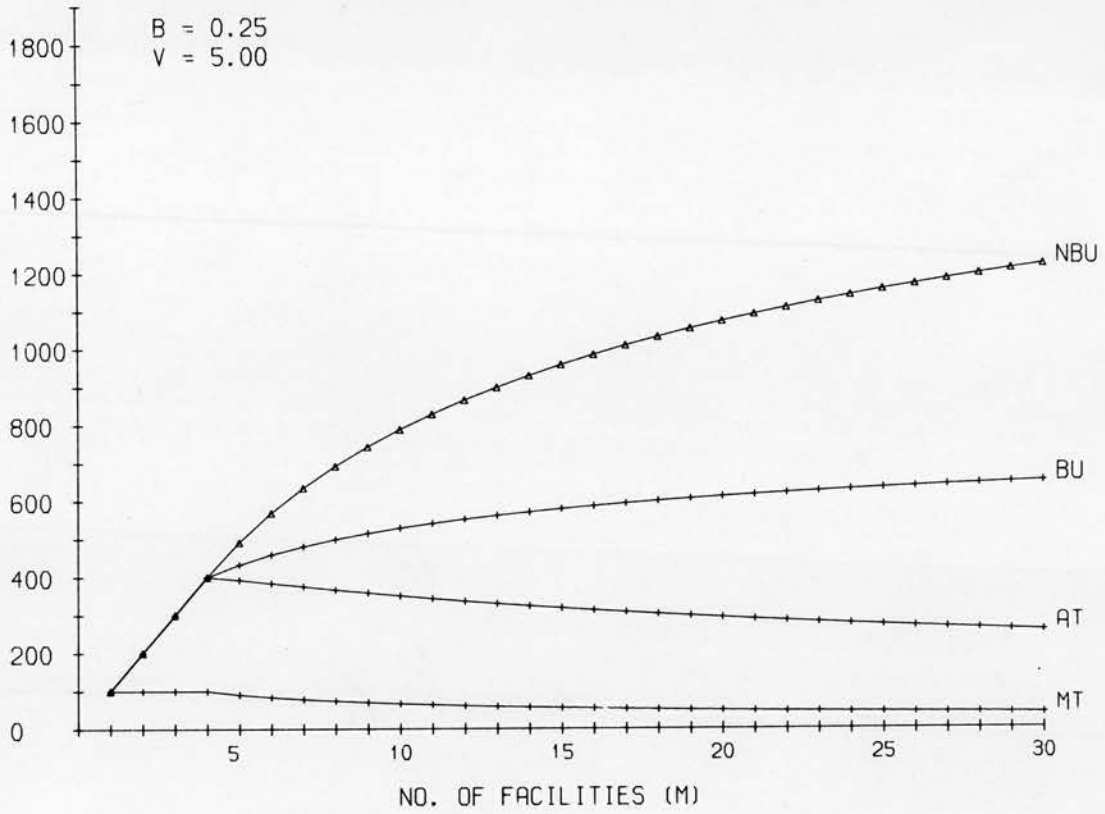


Figure 3.18

Relation between m and NBU, BU, AT and MT, expressed as percentages of values for $m = 1$, on an isotropic plane with $b = .25$ and $v = 5$

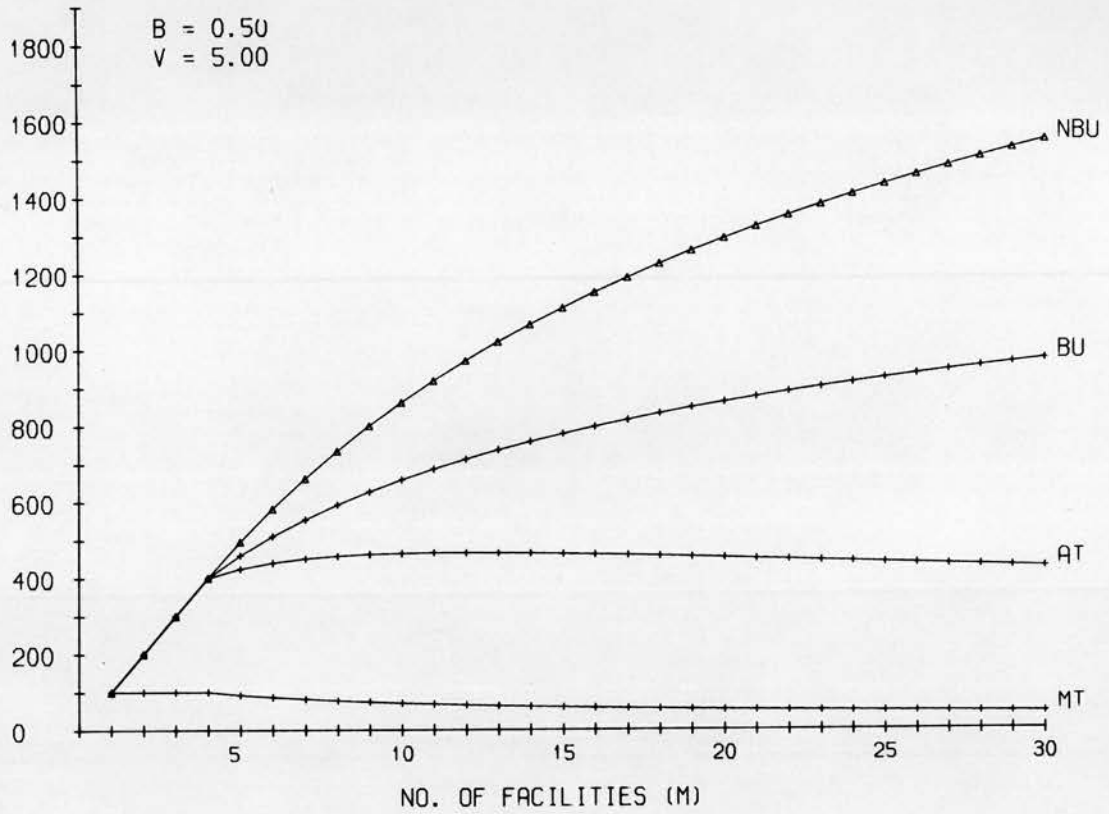


Figure 3.19

Relation between m and NBU, BU, AT and MT, expressed as percentages of values for $m = 1$, on an isotropic plane with $b = .50$ and $v = 5$

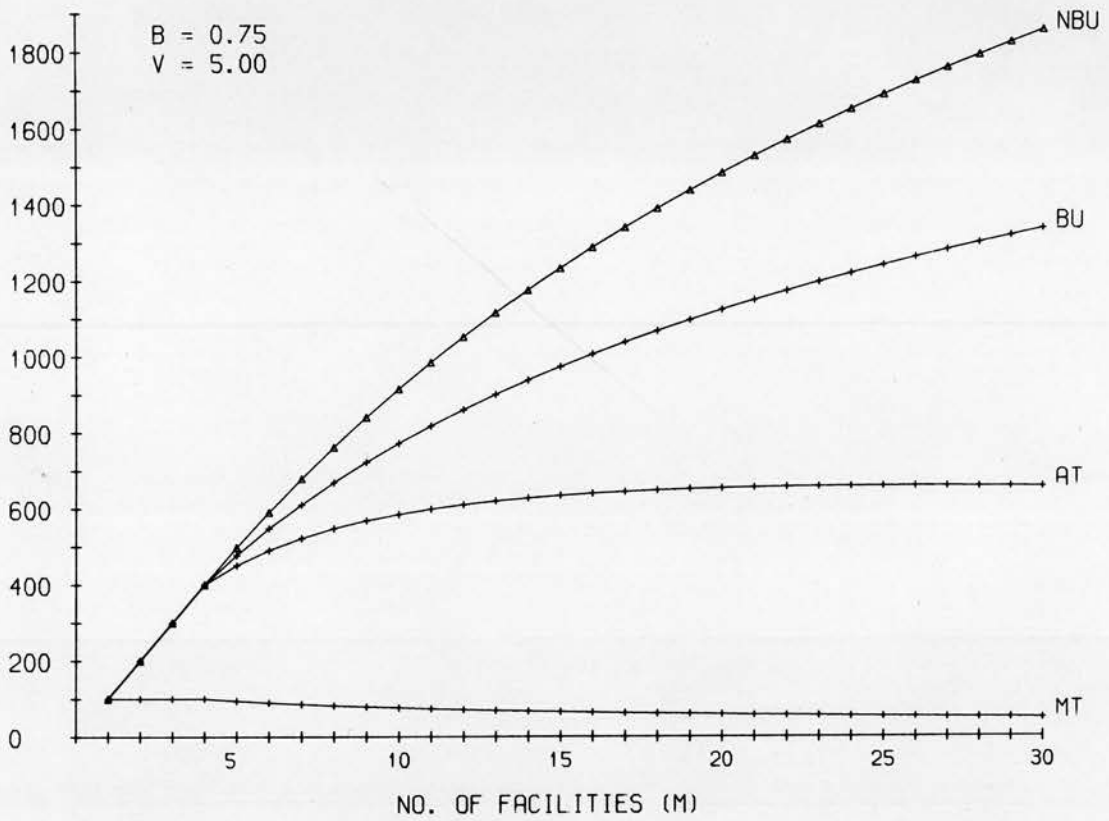


Figure 3.20

Relation between m and NBU, BU, AT and MT, expressed as percentages of values for $m = 1$ on an isotropic plane with $b = .75$ and $v = 5$

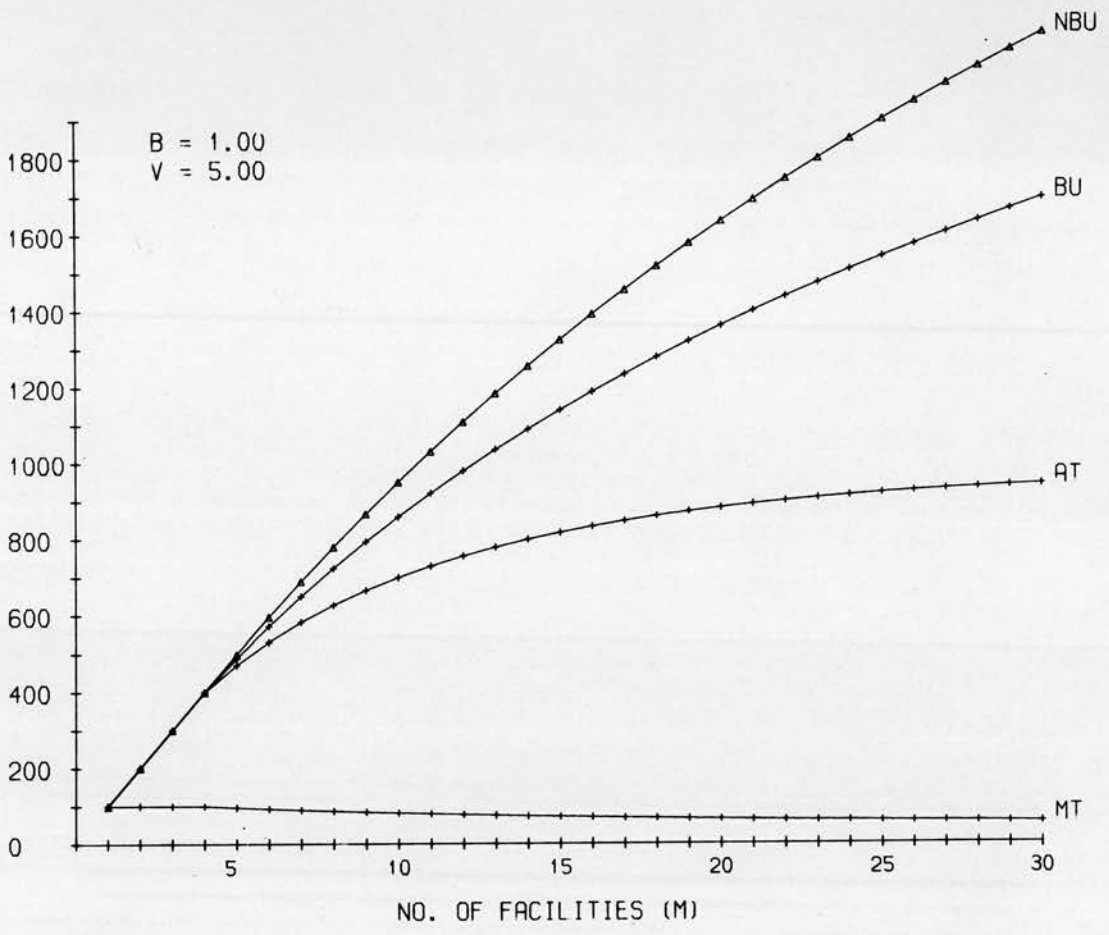


Figure 3.21

Relation between m and NBU, BU, AT and MT, expressed as percentages of values for $m = 1$, on an isotropic plane with $b = 1.0$ and $v = 5$

values of b with $v = 5$. The modification just outlined ensures that BU , AT and therefore NBU rise linearly till $m = 4$; MT is constant to the latter value, declining gently thereafter. Above $m = 4$ the curves have a generally similar form to those for $v = 10$. Again, with $b = 0.25$ and $b = 0.50$ for most of its length the AT curve is falling or nearly level, whereas for $b = 0.75$ and $b = 1.00$ AT rises continually within the range of values shown here for m .

The relative curves for $v = 5$ are generally higher than their counterparts for $v = 10$. At first this appears paradoxical since the value of v is lower. These higher values, however, result from expressing all values relative to those for $m=1$. The values of $TRIPS$, MT and AT are independent of v ; they depend only on b where BU and NBU depend on both v and b . Hence the former indices have the same absolute values for both $v = 10$ and $v = 5$, except in the particular instances when $v = 5$ and m is less than 4. There the absolute values are lower than their counterparts for $v = 10$ because we have now reduced the upper limit of integration in computing $\sum_j T_{ij}$ when v is less than 10. As a result, the demand cones for $m = 1$ have been truncated, so the values of MT , AT , $TRIPS$ and consequently NBU for $m = 1$ are lower than their counterparts for $v = 10$. The higher overall level of the second set of curves is therefore simply a result of expressing the indices as percentages of a lower base.

Isard (1960, pp 525-527) discussed attempts to find a means of integrating population potential and potential transport cost within one framework and concluded that no one had succeeded in resolving this problem. Population potential is in fact a measure of potential demand, like $\sum_j T_j$, while potential transport cost is essentially the same as AT or TC . The framework just presented therefore

resolves Isard's difficulty within the limited context of the location/allocation problem, an unforeseen result of developing it.

Equity

So far we have dealt mainly with efficiency to the exclusion of equity. Equity is harder to define and measure but one way of treating it is to concentrate on the inequality in access, IA , between the nearest and furthest user. In fact this difference is r_{\max} (Fig.3.1). We noted earlier that r_{\max} could be expressed in terms of a or r_{\min} . Thus

$$\begin{aligned} IA = r_{\max} &= (2/\sqrt{3}) a \\ &= (2/\sqrt{3}) a_1 m^{-\frac{1}{2}} \end{aligned} \quad (\text{from 3.1})$$

As m increases IA will therefore fall in a curve which has the same basic shape and a similar gradient to the aggregate travel curve for inelastic demand (3.3). Thus equality of access improves at about the same rate as travel cost falls.

Covering goals

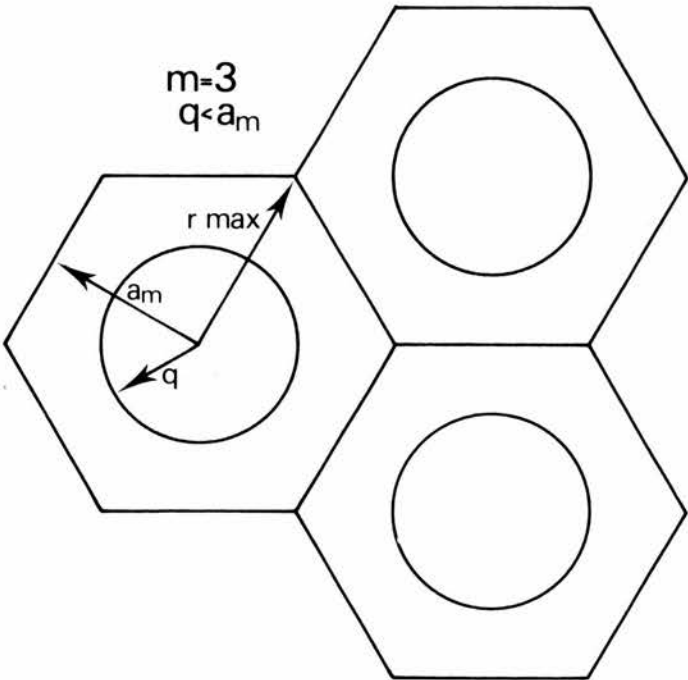
It was argued in Chapter 1 that covering models are relevant to the question of equity. If a satisfactory minimum standard of access has been defined as a distance of q , questions arise of how many centres are needed to 'cover' the whole population and how the actual number covered increases with m . From Figure 3.22a, when $q < a$ the population covered, PC , in one catchment is the population enclosed by a circle of radius q or $\pi q^2 \cdot p$. With m catchments, as long as $q < a_m$, then

$$PC = m \cdot \pi p q^2$$

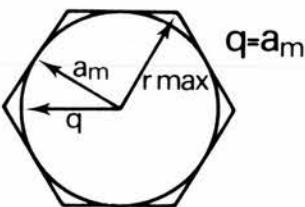
Since $\pi p q^2$ can be regarded as a constant term, PC is linear with respect to m over this range (Figure 3.23).

As m increases, however, eventually a_m will fall to a value less than q . At this point a further increase in m only brings

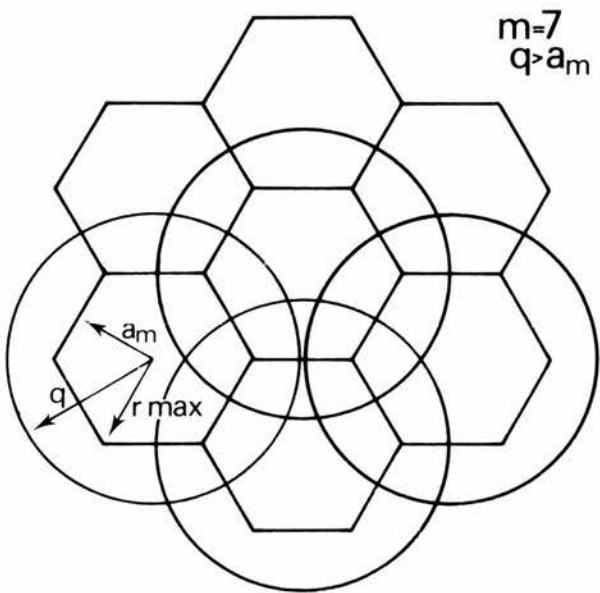
(a)



(b)



(c)



In all of the diagrams above, the hexagons describe catchments;
the circles describe covering radii.

Figure 3.22

Comparison between a constant radius of cover, q ,
and the radius of catchment areas which changes
as m increases

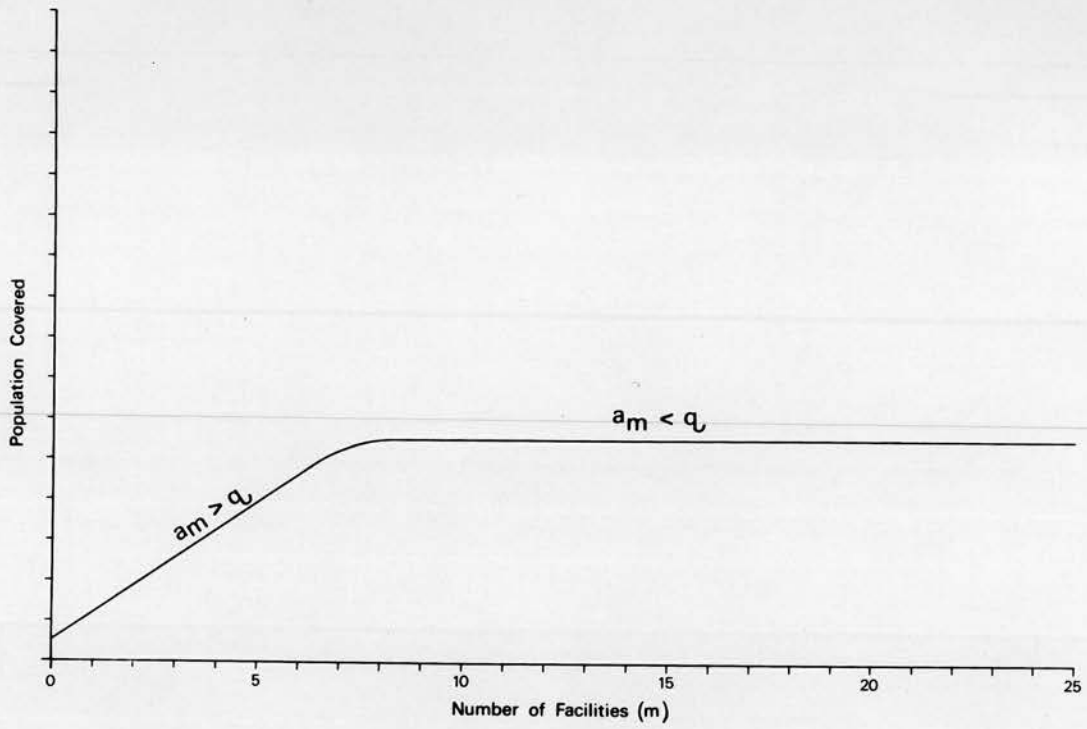


Figure 3.23

Relation between m and population covered

within cover the area within the hexagon but outside the inscribed circle, i.e., the area between $a_m (= r_{min})$ and r_{max} units away (Fig. 3.22b). When m is so large that r_{max} is less than q , the whole population is covered (Fig. 3.22c). Beyond this point the curve of PC lies flat. Of course, the actual shape of the curve will depend on the relative values of q and a , but its general form is now clear enough to allow it to be sketched (Fig. 3.23).

Naturally, when $r_{max} = q$, there are exactly enough facilities to cover the whole population. At this point

$$q = r_{max} = 2/\sqrt{3} a_1 m^{-\frac{1}{2}}.$$

Solving for m

$$\begin{aligned} m^{\frac{1}{2}} q &= 2/\sqrt{3} a_1 \\ \therefore m &= \frac{4}{3} \frac{a_1^2}{q^2} \end{aligned}$$

Hence, given a_1 and the covering radius under consideration, we can determine how many facilities are needed to meet that standard of cover. For instance with $a_1 = 10$ km we would need 34 facilities to give a covering radius of 2 km. To give a cover of 1 km we need 134 facilities.

Conclusion

There is no satisfactory way, at present anyway, of combining equity and efficiency into a single measure. Both are important and both should be considered in plans for locating services. The concept of equity is a particularly complex one, which cannot be adequately treated within the scope of the present work.

Thus we have now devised a series of models which are able to provide simple answers to the initial questions posed about the way in which users benefit from a more dispersed distribution of supply. We have established that the benefits to users of a more dispersed or

less dispersed distribution of centres will depend on whether demand is elastic or not. The nature of demand therefore ought to be taken into account when decisions are being made about the degree to which a service ought to be centralised, a point we will examine more closely in the next chapter.

Our models do not take account of the fact that a few large centres may be relatively more attractive to users. It would be possible to do so by making attraction some kind of inverse function of m , a possible avenue of future exploration.

A crucial thread running through most of this discussion has been the problem of how to develop a satisfactory mathematical model for treating elastic demand and how to integrate this model into a wider framework. This problem turned out to be more difficult than first imagined; in fact, it threw up a number of subsidiary problems. One point emerged clearly from the analysis: the most common formulation of distance decay, the inverse power function has a number of inherent properties, which have not been stressed sufficiently by geographers. If this function is used as a basis for describing elastic demand, at best it has serious idiosyncracies; at worst it produces results which are logically unsound. The negative exponential function does not have such disadvantages and has therefore been used as the basis for locational work in later chapters. Having outlined the various ways in which users benefit from better access to facilities, we can now examine how these benefits compare with the cost of providing more facilities, discussed in the preceding chapter.

CHAPTER 4

The Optimal Number of Centres

Introduction

Previous chapters have discussed the social costs and benefits of making supply more accessible. In a broad sense construction costs and economies of scale argue for a few large facilities; travel costs, elastic demand, equity of access and diseconomies of scale argue for widely dispersed small units. The optimum number of facilities will be determined by the interplay of these factors. In the present chapter an attempt will be made to provide a very simple framework, based on a series of examples, for looking at this interplay. All of the examples below involve the standard assumptions already employed about the distribution of population and supply points.

Example A : inelastic demand

Consider a region or city where entry to a central public service, say swimming pools, is free to users; the service is financed by an annual levy per head of adult population (or per household). This assumption simplifies accounting of costs and benefits but does not affect the outcome of the models. Assume that this flat charge is a full cost price, including all running costs and all capital costs averaged over the lifespan of a pool. A certain size of centre (a pool of 25 metres by 12 metres for instance) has been selected as standard for reasons of convention and/or technical efficiency. The cost of acquiring land is the same for all possible pool locations. The total cost of providing the service, SC , will then be a linear function of the number of centres built. Hence

$$SC = cm$$

where c is the construction cost of one centre.

If we assume for the time being that demand is inelastic, but can be met by a wide range in the number of facilities assigned, we also require the assumption that capacity constraints do not operate over that range, an assumption to which ^{Toregas and}ReVelle (1972) lends some support, as noted in Chapter 1.

In practice, demand for swimming is unlikely to be completely inelastic, of course. Example C below, in which we address elastic demand, is therefore in fact more relevant to swimming pools.

Here the only direct cost to the user of visiting a centre is travel cost. If m facilities are assigned, the aggregate travel (in units of population kilometres) over the whole system AT_m , has already been established:

$$AT_m = k_{AT} m^{-\frac{1}{2}} \quad (3.3)$$

If the travel cost per km for each individual in monetary terms is c_t , this expression represents a total travel cost, TC_m , of

$$TC_m = AT_m c_t = k_{AT} m^{-\frac{1}{2}} c_t \quad (4.1)$$

The greater the number of centres, the greater will be the savings in the travel costs of users. Since demand is inelastic we must have at least one facility to meet the needs of users. Everyone in the landscape would then have to use that facility, resulting in a certain overall transport cost, TC_1 . Consider now the possibility of building two centres instead of one, thereby reducing overall travel cost to TC_2 . The latter situation represents a savings in travel costs of $TC_1 - TC_2$, denoted by TS_2 . Similarly three facilities, as opposed to the basic requirement of one, would represent savings in travel of $TC_1 - TC_3$. Thus in general

$$TS_m = TC_1 - TC_m \quad (4.2)$$

In fact we already have an expression for TC_m in 4.1; thus where

TS_m is the savings in travel accruing from m facilities,

$$TS_m = TC_1 - k_{AT} m^{-\frac{1}{2}} c_t \quad (4.3)$$

In this expression TC_1 is a constant term with respect to m .

Thus we have used the travel costs when $m=1$, as a basis for defining the travel savings obtained from two or more facilities.

In this framework by definition TS_m must be zero when $m=1$. Hence if we substitute $m=1$ and $TS_m = 0$ in 4.3 we have

$$\begin{aligned} TC_1 - k_{AT} c_t &= 0 \\ \text{i.e. } TC_1 &= k_{AT} c_t \quad (4.4) \\ \text{i.e. } AT_1 c_t &= k_{AT} c_t \\ \text{i.e. } k_{AT} &= AT_1 \end{aligned}$$

We can now use 4.4 to rewrite 4.3

$$TS_m = k_{AT} c_t - k_{AT} m^{-\frac{1}{2}} c_t \quad (4.5)$$

By defining an appropriate standard value for the constant c_t we can sketch the form of the TS curve (Fig.4.1). As before, we take $c_t = 1$.

It is implicit in 4.2, that the amount by which TC falls when m is increased by one at any stage, is the amount by which TS rises. Hence the curve of travel savings is simply a mirror image of the curve for travel costs. It may be worth pointing out here that the TC and TS curves both give absolute values; the marginal cost or benefit for any increment in m has to be calculated from the gradient of the appropriate curve at that point.

Let us assume that the city wishes to determine the best number of

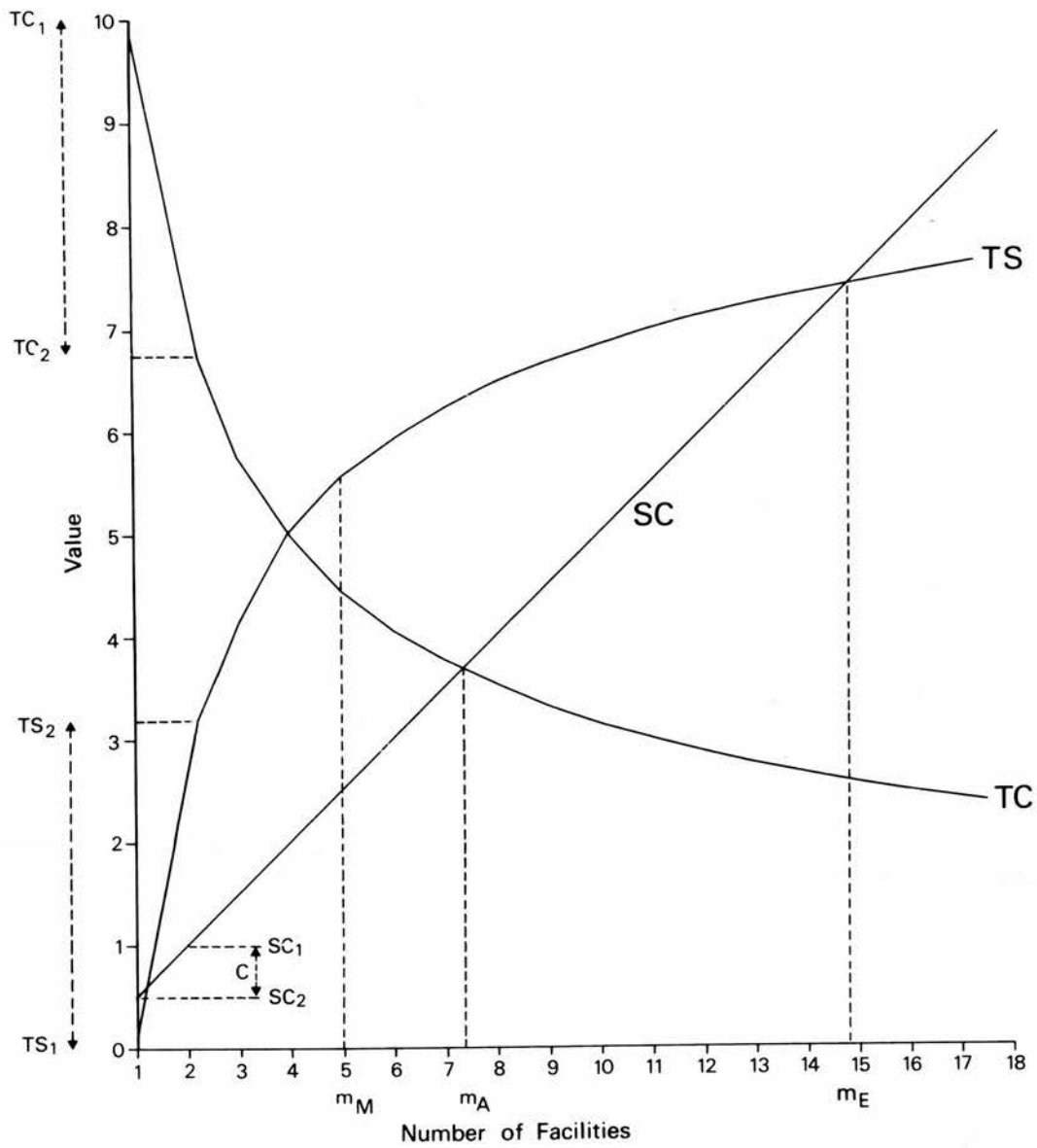


Figure 4.1

Relation between m , savings in travel (TS) and cost of supply (SC)

pools to provide in the long term. Citizens will be aware that by constructing more pools, they incur a higher levy but reduce their travel costs and vice versa. In weighing up these two factors, two philosophies can be invoked: one based on marginal, the other on total benefit. Under the first, pools should be built to the point where marginal gain just exceeds the marginal cost. According to the second the city should try to "break even", constructing pools to the point where the total benefit in travel savings just exceeds the total cost of construction.

If the marginal approach is applied to the situation sketched in Figure 4.1, the reduction in travel costs brought about by building a second pool is $TC_1 - TC_2$ (or $TS_2 - TS_1$), which exceeds the cost of construction ($SC_2 - SC_1 = c$). As construction of this pool leaves the city's population better off, in a general sense, it represents a worthwhile investment.

However, since TC tends to level out, at higher values of m the marginal savings in travel will eventually be less than the construction cost; net marginal benefit is then zero or negative. It will therefore pay in the long term to build pools up to the point where the marginal gain in travel savings just exceeds the construction costs; this is the point at which the gradients of TS and SC (with respect to m) are equal. Obviously the gradient of SC is c . Differentiating 4.5, the gradient of TS is

$$\frac{d(TS)}{dm} = + \frac{1}{2} k_{AT} m^{-1.5} \cdot c_t$$

The optimal value of m is therefore given by

$$\begin{aligned} c &= \frac{1}{2} k_{AT} m^{-1.5} c_t \\ \text{i.e.} \quad 2c(k_{AT} c_t) &= m^{-1.5} \\ \text{i.e.} \quad m &= \frac{(k_{AT} \cdot c_t)^{2/3}}{2c} \end{aligned}$$

In a brief discussion of this question, Abler, Adams and Gould (1971, p. 549), apparently following Törnqvist (1963), argue that the optimum is where the SC and TC curves intersect. Now this intersection occurs where

$$c_m = k_{AT} m^{-\frac{1}{2}} \cdot c_t$$

which yields

$$m = \left(\frac{k_{AT} c_t}{c} \right)^{2/3}$$

This value of m is approximately 1.59 times larger than the previous one. The optimal value of m from a marginal viewpoint is therefore smaller than that suggested by Abler et al. Therefore Abler and co-authors must either have had an approach other than the marginal one in mind or they are simply in error.

Applying the 'break even' approach to the same situation would involve selecting that value of m where travel savings just exceeded or equalled the total cost of supply. In fact, TS and SC cut at two points. The lower value, very near the origin, can be regarded as a trivial solution and will be ignored henceforth. A break even approach therefore leads to m_E (Fig.4.1) pools being constructed - a somewhat larger value than the previous optimum. The nature of the TS and TC curves suggests that the break even point will usually be larger than the marginal optimum.

In so far as any financial goal analogous to profit and loss is appropriate for a public agency, especially if one providing essential services is being considered, the break even approach seems the more appropriate of the two; for instance, it is the philosophy recommended to the nationalised industries in United Kingdom by those acts of Parliament which brought them into being. Furthermore, since it leads to a larger value of m - and therefore a better spatial spread of supply - it provides a more equitable spatial distribution. On the other hand, for a problem in the private sector (e.g., involving the

location of warehouses), the marginal solution would probably be of greater interest. Since Abler et al. argue that their solution is where transport savings balance costs of production, it is possible that they had a break even approach in mind, though unstated. However, their solution, m_A (Fig.4.1), is quite different from that optimum as well; in fact, SC and TS must always intersect at a different point from SC and TC, apart from one special case, the point where TS and TC cut. Whichever approach they had in mind, there would therefore appear to be an error in their analysis. Although they probably only intended to illustrate the nature of the problem, these errors may be worth pointing out since their treatment is quite a popular one; in fact, it provided a helpful starting point for the present discussion.

The preceding analysis assumes that m is a continuous variable. The fact that it can have only integer values is not a serious difficulty since we can take the nearest integer value as a satisfactory approximation.

Example B : economies of scale

We can now take account of possible economies of scale. Suppose size of facility is no longer fixed but depends on m so that when m is small, we have a few large facilities, each serving a wide area, and when m is large, the facilities and their catchments are small. This is implicitly true in the previous example since the capacity of each facility will depend on the size of its catchment and therefore on m . If there are consistent economies of scale in construction, SC will be concave upwards (SC^E in Fig. 4.2); with diseconomies it will be concave downwards (SC^D). The SC^D curve, showing diseconomies, could actually have various positions, depending on whether the diseconomies set in early with fairly low values or later with fairly high

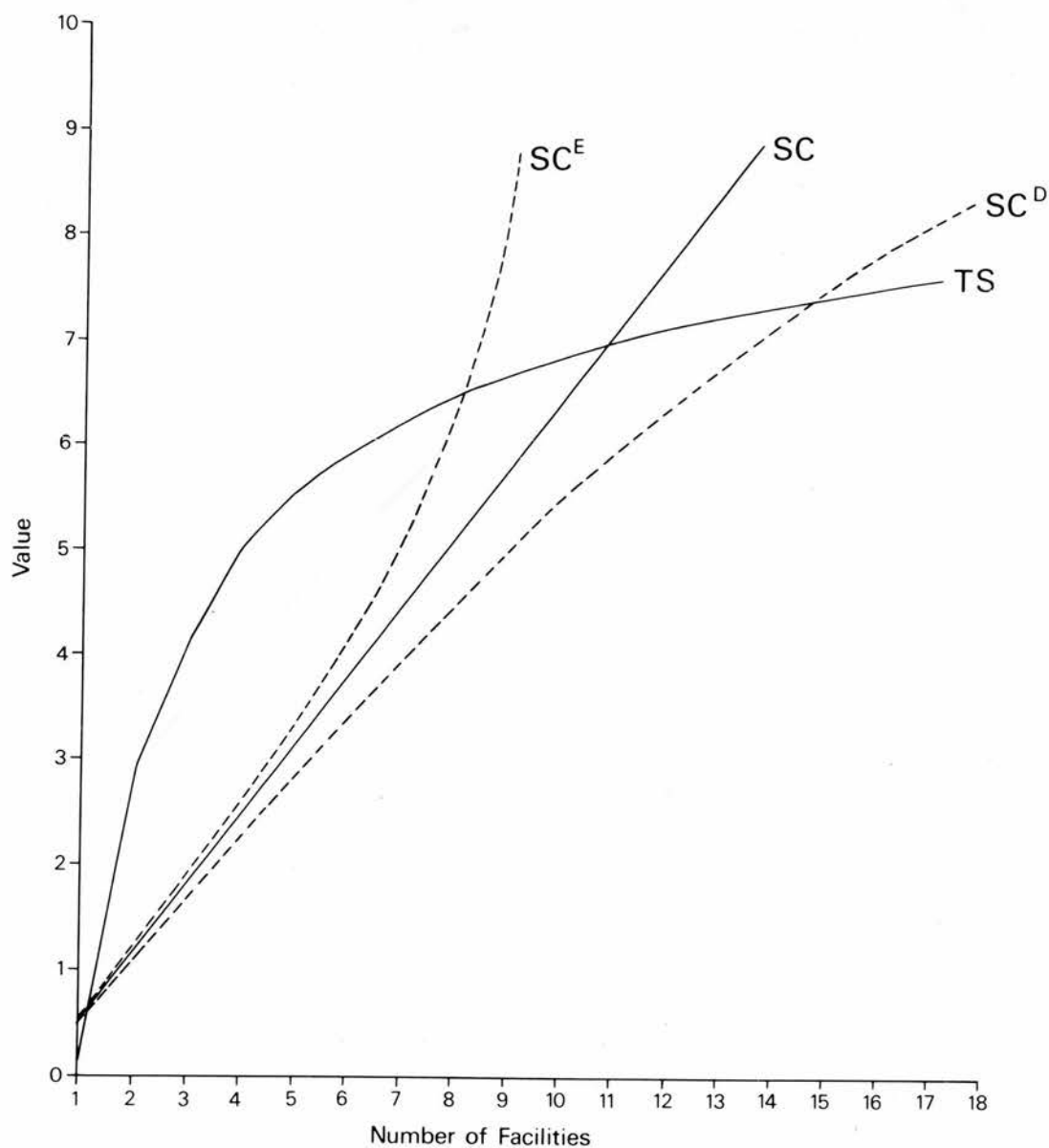


Figure 4.2

Relation between m , savings in travel (TS) and cost of supply (SC) illustrating economies (SC^E) and dis-economies (SC^D) of scale

values of m . In view of the general conclusions of Chapter 2, it seems more reasonable to assume that they will set in fairly gently at first but somewhat more strongly with higher values of m , as sketched in Figure 4.2.

In these circumstances the point where the gradients of SC^E and TS are equal will now be to the left of its previous position. Similarly the point where TS breaks even will now be further left. Conversely, with diseconomies the equivalent point will be to the right of their former values. Economies of scale therefore tend to lower the optimal value of m ; diseconomies tend to raise it. With more complex SC^E and SC^D curves, however, the conclusions might conceivably be different.

Example C : elastic demand

If demand is elastic, we can argue that it will be worth building centres as long as the marginal net benefit to users, NBU , exceeds marginal cost or total benefit exceeds total cost. The latter problem turns out to be the easier one so we will examine it first.

Figure 4.3 shows SC in relation to the absolute value of NBU for different types of cone ranging from the gentle ($b = .25$) to the relatively steep ($b = 1.0$). Clearly, the break even point for a gentle cone will always exceed that for a steep cone, irrespective of whether the gradient of SC is steeper or not. The reason is simply that the lower value of b represents a higher overall level of demand. As a corollary, the optimal value of m with inelastic demand will be even higher than that for $b = .25$. Yet, an opposite conclusion can be reached in two different kinds of circumstance. As a prelude to the first of these exceptions it is worth recalling here that demand cones can differ in the intercept values, f_0 , as well as their slope values, b . The above comparisons were based on the assumption made earlier

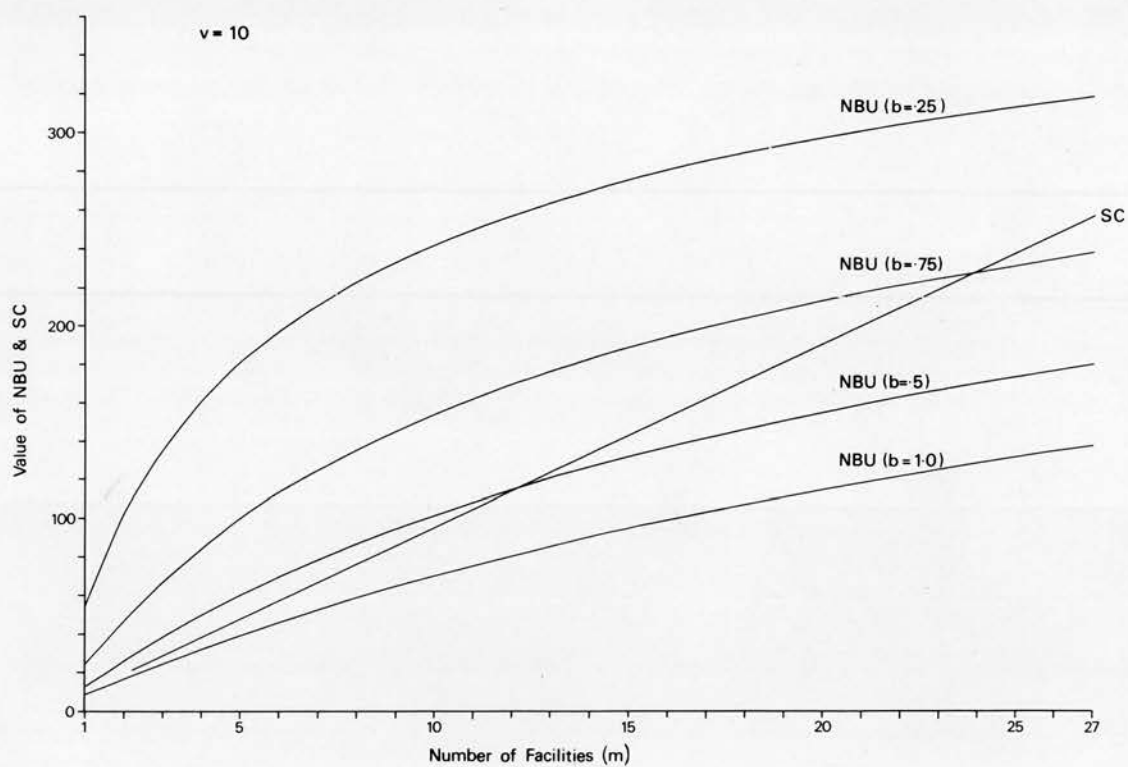


Figure 4.3

Relation between m , net benefit to users (NBU) and cost of supply (SC) for selected values of b

that all the cones concerned have the same intercept values. Suppose a local authority is considering whether to build more indoor tennis courts or more squash courts, given that a few of both are in existence and are quite well located. Let us also suppose that the demand cone around one squash court is steeper, say $b = 0.5$ but has a much higher intercept with the result that the volume of demand under it is the same as that under the demand cone for indoor tennis ($b = 0.25$). Stated more plainly, for squash there are potentially more adherents but participation falls off more rapidly with distance. Since total demand for both activities is the same when $m=1$, the absolute values of demand and net benefit can be projected by referring to the relative response curves discussed earlier. In consequence the benefits of building more squash courts will therefore be greater in absolute as well as relative terms. Thus the optimal value of m will, in this instance, be larger for the activity with the steeper cone. Given a certain budget to divide between the two, other things being equal, squash should have the larger share and its courts should have a more dispersed distribution. Although this example is a consequence of contriving, somewhat artificially, an f_0 value which exactly offsets a higher b value, it makes two useful points. First, intercepts matter as well as slopes. Second, the fact that equal volumes of demand are obtained at one centre for two services, does not mean that their supply points should have the same distribution.

Of necessity the discussion has assumed so far that all facilities are well located and that each part of the city is therefore relatively well provided. The second exception involves relaxing this assumption for a moment. Let us suppose that most of the city's recent growth has been in one direction and that as a result, a large area on the western side of the city lacks many public services, including say

public libraries. Now, if demand for libraries is inelastic, construction of libraries on the western side or anywhere else will not increase the overall utilisation of the service. Such provision will, of course, reduce aggregate travel in the city as a whole and in particular for readers near the new libraries. It will also divert some demand from older libraries adjoining the new area and thereby reduce their level of use. In contrast, if demand is moderately elastic, a substantial number of new readers will be added as well as mean travel being reduced for existing readers. Furthermore, the loss of readers from existing libraries will be smaller. On balance, therefore, if part of the city is relatively underprovided a stronger case for expanding supply could conceivably be made in absolute terms with the values used in the existing framework, when demand is moderately elastic. Of course, the absolute value of benefit from a new centre could still be greater with inelastic demand than with elastic demand simply because the term concerned with evaluating the reduction in travel cost (3.16) will affect more people.

Hitherto the value of newly created demand has been measured on the same scale as reduction in travel cost to existing users. However, making a value judgment outside the present framework, we could argue that 'new' demand merits extra weight because it represents both a broader level of participation and a more equitable spatial distribution of opportunities. Thus equity strengthens the case for a spatial extension of supply when demand is elastic. Perhaps it is worth noting here how important it is to know the elasticity of demand when trying to forecast the impact of a new facility in an underprovided area of the city.

It may be argued that demand for all services depends on the social attributes of an area's population: age, social class, income and so forth.

Though this is true, the point remains that in each of the categories of population formed by disaggregation, demand is either inelastic or elastic with respect to travel cost or distance. The argument developed here will therefore apply within each of those categories. Hence, although the outcome in relation to the present example will be more complex because we then have to balance gains to different social groups, the need to disaggregate does not essentially invalidate the structure of the existing model.

The preceding discussion of example C has been concerned with the break even point. As noted at the start, finding the 'marginal' optimum seems to be an even more puzzling question. Table 4.1 shows the values of NBU and the increments added by successive facilities as m increases for the standard values of the relevant constants with $b = 0.25$. Similar tables could be presented for other values of b . To obtain the marginal optimum it is helpful to define a specific increment in benefit or use which an extra centre must add to be viable; this 'threshold' value can be expressed in terms of marginal net benefit to users (NBU) or simply in terms of the number of extra trips the centre generates. Clearly, for our purpose here the threshold value should be defined so that it is equivalent to the cost of supplying an extra centre. We can argue that when the increment in NBU or in demand falls to a level equal to the threshold, we have reached the optimal level of m .

With a relatively high threshold, say 10 units of NBU, and $v = 10$, Table 4.2a shows that 9, 8, 5 and 0 facilities would be worth building respectively as we go from a gentle cone ($b = .25$) towards the progressively steeper cones ($b = .50, .75$ and 1.0). This result is consistent with the previous conclusion that the more elastic the demand the lower the optimal value of m . However, if

we take the lower threshold of 2.5 then the respective optimal values of m are 28, 33, 34 and 33, almost reversing the previous pattern completely.

Table 4.1 Response of TRIPS, MT and NBU to increase in
 m from $m = 1$ to $m = 30$ with $b = 0.25$ and $v = 10$

<u>m</u>	<u>TRIPS</u>	<u>MT</u>	<u>NBU</u>	<u>INCREMENT to NBU</u>
1	11.40	5.1	55.6	55.6
2	16.88	3.9	102.0	46.4
3	20.30	3.3	135.0	33.0
4	22.74	2.9	160.0	25.0
5	24.60	2.6	179.9	19.8
6	26.09	2.4	196.3	16.3
7	27.32	2.3	210.1	13.7
8	28.36	2.1	222.0	11.8
9	29.26	2.0	232.3	10.3
10	30.05	1.9	241.5	9.1
11	30.74	1.8	249.7	8.1
12	31.36	1.8	257.1	7.3
13	31.92	1.7	263.8	6.7
14	32.43	1.6	269.9	6.1
15	32.90	1.6	275.5	5.6
16	33.33	1.5	280.7	5.1
17	33.72	1.5	285.5	4.8
18	34.09	1.4	290.0	4.4
19	34.43	1.4	294.2	4.1
20	34.75	1.4	298.2	3.9
21	35.05	1.3	301.9	3.6
22	35.33	1.3	305.3	3.4
23	35.59	1.3	308.6	3.2
24	35.84	1.3	311.8	3.1
25	36.08	1.2	314.7	2.9
26	36.30	1.2	317.5	2.8
27	36.51	1.2	320.2	2.6
28	36.72	1.2	322.8	2.5
29	36.91	1.1	325.2	2.4
30	37.10	1.1	327.6	2.3

Table 4.2 Optimal values of m in a marginal framework for
given thresholds of NBU

(a) With $v = 10$

<u>Threshold in units</u> <u>of NBU</u>	<u>Optimal value of m</u>			
	<u>$b = .25$</u>	<u>.50</u>	<u>.75</u>	<u>1.0</u>
30	3	0	0	0
20	4	3	0	0
10	9	8	5	0
5	16	17	16	12
2.5	28	33	34	33
1.5	42	51	56	57

(b) With $v = 5.0$

<u>Threshold in units</u> <u>of NBU</u>	<u>Optimal value of m</u>			
	<u>$b = .25$</u>	<u>.50</u>	<u>.75</u>	<u>1.0</u>
20	0	0	0	0
10	5	0	0	0
5	10	8	0	0
2.5	18	18	15	10
1.5	28	30	29	25
1.0	39	43	44	42

Since the value of NBU depends more on v (the value we assign to each trip) than on any other parameter, it is interesting to compare these results for the relatively high value of 10 with those for the lower value used before, $v=5$. In general, a similar pattern is found, as Table 4.2b shows. Thus, for such high threshold values of NBU as 20, 10 and even 5, the optimal value of M falls as demand becomes more elastic. At the lower threshold values of 1.5 and 1.0, however, the values for $b = .50$ and $b = .75$ slightly exceed that for $b = .25$, reinforcing the previous paradox.

Perhaps just as significant is the tendency in both halves of the table for b to make less and less difference to the optimal value of m as we move to lower threshold values. On the surface this is surprising because there is a very great difference between the volume of demand under cones defined by $b = .25$ and $b = 1.00$. We might therefore have expected this difference to be reflected in the values of m at low as well as high thresholds.

Adhering to marginal logic can therefore lead to somewhat paradoxical conclusions since, depending on the cost of constructing a centre, the optimal level of supply can be greater for elastic demand than for inelastic demand, despite the fact that the absolute amount of demand will always be greater in the latter case.

To summarise, there may be circumstances in which elastic demand leads to a higher optimal value of m rather than the lower value generally expected. Yet it is also tempting to regard this paradoxical result with ambivalence; in a sense it is a hazard, even a defect, of the marginal approach. This is implicitly a mild criticism of Christaller and L sch since their models are based only on a marginal approach. Since our landscape is similar to L sch's, in a roundabout

way we have already established that a break even approach can be developed for a Löschian landscape (Figure 4.3).

Example D : Private Suppliers

It is interesting to consider what the outcome might be were the service to be provided by private entrepreneurs. The entrepreneur has no reason to consider the travel costs of users: if he imposes a flat charge for admission and his costs are fixed, he will simply try to locate so as to maximise the amount of demand he draws. If demand is elastic, competing suppliers will find it advantageous to locate in a dispersed manner, as in Christaller's model. Conversely if demand is inelastic, the advantages of dispersed locations will appear to be less tangible to the suppliers themselves; as a result they may tend to cluster near the first centre (as in Hotelling's case) or near the centre of the city or may form loosely dispersed clusters. Thus a competitive environment could force a more centralised pattern of supply - to the disadvantage of users, since the social optimum would still be uniform dispersion. A uniform distribution of supply points can no longer be taken for granted as a basic element in the model with fixed demand.

If the threshold value for an entrepreneur is the same as in Example C, the optimal values of m for inelastic demand will be potentially the same, because the amount of total demand remains unchanged. However, the actual number may depend on the outcome of the locational strategies of suppliers; as we have just noted, this outcome is uncertain. With elastic demand suppliers will still be uniformly spaced, so that the private optimum and social optimum coincide spatially. Given the same threshold, the optimal values of m will therefore be the same. Thus, within the confines of our model,

private provision of the service could lead to a different spatial arrangement and possibly a slightly lower level of supply with inelastic demand, but not with elastic demand.

If economies or diseconomies of scale are introduced, the effect will be the same as before: a reduction and increase, respectively, in the optimal value of m . This applies to private as well as public suppliers and to elastic and inelastic demand alike.

Example E : A Spatially Flexible Supply of Given Amount

In all the examples discussed so far supply has consisted of a variable number of units of a standard size or capacity. Suppose instead that total supply is fixed but that centres can vary fairly freely in size. For instance, public libraries might range from those with 100,000 books and three or four staff to those with 10,000 books, open for only half the day. Suppose that within a certain range in size, costs are such that 30 'half time' libraries cost the same as 15 full time ones with a single librarian or 5 large enough to need three librarians. Under these conditions we can pose the question of how to organise a given amount of capacity. SC is now fixed and therefore lies flat parallel to the x -axis (Fig.4.4); in other words we can purchase higher values of m at no extra cost in construction.

For a service of given elasticity, say $b = 0.25$, the community can break even by picking any value of m where the gains of users exceed the costs of supply, i.e., where NBU lies above SC . If we define net social benefit, NSB , as the margin by which NBU exceeds SC , then we can maximise NSB by moving rightwards along the curve for $b = 0.25$ towards higher levels of NSB . Thus seven facilities clearly yield a higher level of net social benefit than six in Figure 4.4 in marginal terms. For all values of b within the

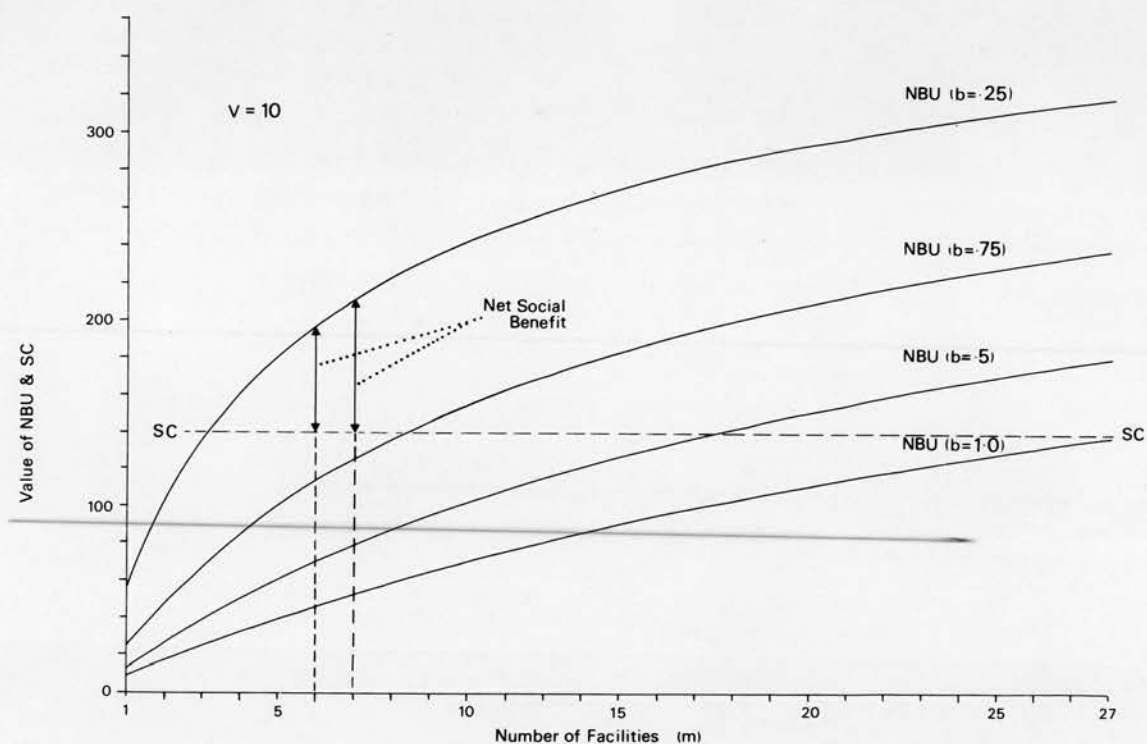


Figure 4.4

Relation between m , net benefit to users (NBU) and cost of supply (SC) when supply is spatially flexible. (Example E)

present model it will then be better to opt for the largest feasible value of m within the range where SC is horizontal. In these circumstances, a solution with many small centres combines efficiency and equity in a most attractive way.

In reality this exact situation is unlikely, because there are certain indivisible amounts of capital needed for almost all services. However, some services may well have a certain latitude over which m can be increased and size reduced proportionately with little impact on total cost. Such conditions are most likely to be found in services where labour is the major cost, such as schools and libraries. Elastic demand is then a strong argument for choosing the smallest size within that range; the extra demand, participation and reduced travel costs come as virtually free bonuses. The fact that users then enjoy a more equitable level of access to the service is also worth stressing. A recent letter to the 'Times' by T.R. Lee (Sept. 5th, 1978) would seem to suggest that the argument just outlined has some relevance to primary and secondary education, though demand for these services is not generally taken to be spatially elastic, whatever attendance records might reveal. The less tangible, qualitative advantages of a smaller scale are also worth recalling in this context.

Naive as it is, this would seem to be one of the more valuable insights which the preceding models have helped to highlight. Indeed, a worthwhile research project could be undertaken to identify which services are sufficiently flexible in terms of the spatial organisation of their supply capacity to enjoy the kind of latitude just outlined. Despite its relevance to such issues as the provision of services in country areas, particularly education, the author is not aware of any work which distinguishes between spatially flexible and spatially rigid

types of service as has been done here. Nor have any attempts been found to probe the problem empirically in the geographical or related literature. It may be worth noting that a useful contribution could also be made on the conceptual side by developing a framework for treating supply conditions which would be general enough to include both the rigid case (Examples A and B) and the flexible case (Example E) as the opposite extremes of a broad continuum of types of facility. The latter task, however, lies outside the scope of the present work.

While giving qualified support to the notion that 'small is beautiful', for this type of service, we have to bear in mind the possibility that for other services larger facilities are more attractive to users because, amongst other reasons, they offer extra amenities. As noted earlier, a simple index of attractiveness could be incorporated into a spatial interaction model to help describe the preferences of users for centres of varying attraction. There would therefore appear to be no theoretical reason why the preceding models could not be extended to include a further range of behavioural parameters. Since we would then have facilities which differed in attraction, this would involve sacrificing the framework of an isotropic plane, a construct indispensable to the argument of the previous two chapters.

Conclusion from Examples

In summarising this discussion, it is worth recalling that we have used m as a single parameter to define the spatial organisation of supply. As a general conclusion we can say that when demand is inelastic the optimal value of m depends on whether we wish to break even or maximise 'social profit' in the form of total net benefit. When demand is elastic the extra parameters involved (b and to some

extent f_0) create a more complex situation. If supply cost is linear with respect to m , on the whole fewer facilities will generally be viable the more elastic the demand. In a real situation with an uneven distribution of population the decision whether to build one extra centre or two smaller ones or none at all should depend on b , f_0 , supply cost factors in form of the threshold, selection of a break even approach or a marginal one, the disposition of existing centres with respect to population and the balance struck between equity and efficiency.

This conclusion and the relationships which buttress it are essentially independent of the values chosen for the various constants: $a_1, k_1, \dots, k_4, k_{AT}, v$ and so forth. With different values for these constants the optimal values of m in a given context may of course change but the structure of the argument and the general relationships leading to that specific result will not alter. Arbitrary values had to be chosen for the constants mainly so that we could illustrate the relationships graphically.

With the wisdom of hindsight we can say that the way in which the optimal value of m was found to depend on the elasticity of demand should have been expected from common sense. In fact, before the analysis was undertaken this relation was only expected in a rather vague, ill defined manner. The mathematical form of this dependence, however, could not have been guessed in advance. Nor could the various exceptions, surprising and puzzling as some were.

Little work has been published on the interaction between the spatial organisation of a service and the response of users; the field is especially lacking in theory. By making a distinction between fixed and elastic demand and by integrating user response

and travel cost in the wider framework of net benefit, we have created a series of models with which we can address this question. However simplistic the framework and models may be, seen against the background of the existing geographical literature on public services, they are probably the major contribution made in the present work.

'Social science is just common sense with big words' a physics student once remarked. Perhaps locational modelling is just common sense dressed up in mathematical symbols. Even so, a clear framework in which the vital assumptions are explicit may help to sharpen common sense arguments, stimulate fresh insights and expose problems which need further investigation.

No doubt planners and entrepreneurs will already be conscious of those factors most relevant to their own needs and interests. In fact existing social practices may well provide very sensible solutions to many problems which are difficult to solve in theory. Whether this is so or not remains to be determined, but it seems reasonable to assume that not all the actors will be conscious of all the costs and benefits or how they effect society as a whole. For this reason, despite all its obvious shortcomings, models of the type developed here may have some useful role in broadening the understanding of the problem, if only because they outline the primary relationships between some of the main factors involved. The fact that the argument is both intrinsically spatial and mathematically explicit may be a further advantage.

SECTION III

Methods of locating centres

CHAPTER 5

Exact methods of solving the location problem in a network

Introduction

Three primary elements can be varied in designing a system of service points: number, location and capacity. The previous section was concerned only with the number of centres. By considering the spatial interplay of supply and demand it was possible to outline a general framework within which the optimal level of supply could be examined. During that discussion it was necessary to set aside the question of finding optimal locations for a specified number of centres. The problem of locating centres, however, is the main concern of the present section and will be the focus of all the remaining discussion. In this section the emphasis is on concepts and methods relevant to the location problem; applied aspects of the location problem will be examined in a later section.

Forms of the location problem

Whether it is formulated as a cost minimization, use maximization or covering problem, three fundamental spatial forms of the location problem can be identified:

1. assign m facilities freely, i.e., assume no facilities already exist in the area (the general problem);
2. locate k additional facilities, taking the existing centres into account (the additional or incremental facility problem);
3. given m existing centres, reorganise the system by closing any badly located centres and allowing a certain number to be opened (the reorganisation problem).

Most writers have concentrated on the first problem, perhaps because the second appears to be easier computationally; the third has received very little attention but can be solved by adapting methods for the general problem. Except in the special cases of an entirely new service or of a completely new town or in the unlikely event of all existing capital being written off, the second problem is much more likely to occur in reality than the first. This is worth emphasising because it has one idiosyncrasy which arises when certain standard methods for solving the general problem are applied to it - a difficulty which will be discussed later. Considerable progress has been made over the last decade in solving the general problem when the objective is to minimise aggregate travel with fixed demand. As noted earlier, much less work has been done with other objectives.

Geometric frameworks

The choice between a network and plane is basically a choice between different sets of assumptions regarding travel by users; the geometry selected determines in turn which methods of space searching can be used to find the optimal solution. If movement is free to take paths which approximate to straight lines or if the network of roads or railways is so dense that relatively direct paths are the rule then the main assumptions of a plane are satisfied. If these rather demanding assumptions are not met and movement is restricted to certain routes or channels, the geometry of a graph is more appropriate; travel cost can then be measured as the shortest path in time or distance through the network. Normally we assume that travel cost is a simple function of distance whereas in reality it is likely to be a more complex function of distance. The degree of approximation implicit in this assumption of course depends on particular circumstances.

The network framework is clearly the more general case; in fact a plane can be viewed as a graph which has the special property that all points have straight line links to all other points. Where there are few topographic barriers, where many people walk to the central point and there is a high density of routes, Euclidean distance may well give a good approximation to travel time in a network. For handling a plane with topographic barriers the LAP algorithm by Goodchild (Rushton et al., 1973) has the useful feature of allowing distances on a plane to be measured round obstacles, which somewhat mitigates the rigidity of this geometry.

The solution of the p -median problem on a graph has been made easier by a theorem proved by Hakimi (1964 and 1965):

"There is a set of p points, consisting entirely of nodes of the graph, which minimises the sum of the weighted distances to the closest of any p points on the graph. (However, another set of p points, not all nodes, could possibly provide the same minimum)."

This statement of the theorem is due to ReVelle, Marks and Liebman (1970), who note that it has been extended by Levy (1967). To find a single median or several medians it follows from this theorem that only the nodes need to be considered, which simplifies the searching procedure. In any case, if facilities are being assigned to dispersed urban nodes within a region, the locations along the routes between the nodes will not normally be of interest because they lack the infrastructure of towns. An interesting aspect of the theorems of Hakimi and Levy is that they seem to be among the few pieces of location theory in existence for networks.

The extra assumptions inherent in the plane purchase two advantages: first, distance can be computed directly by the Euclidean or city block metric, avoiding the need for a shortest path algorithm; second, and more important, the gradient of the objective function at any point may be obtained by using differential calculus, which facilitates search.

In a strict sense, a searching strategy should respect the intrinsic properties of the space used. Since it is not a continuous space the notion of a gradient is inappropriate for a graph. However, in the practical art of devising heuristics to obtain approximate solutions to large problems, it may not always be helpful to obey this rule too rigidly. Thus a useful procedure for examining locations on the road network of an indented coast (Robertson, 1974 and 1976) makes use of Törnqvist's "hill-climbing" strategy, thereby applying a gradient type of search to a graph, an approach which is not strictly valid. On the other hand, since a plane can be viewed as a special case of a graph, methods of searching a network could be applied on a plane by treating the points of a lattice or grid as nodes. This will usually be very inefficient because it does not exploit all the information inherent in the geometry of a plane. As an aside, it is interesting to note that traditional location theory invariably assumes the restrictive geometry of a plane. Algorithms for locating facilities could therefore be used to translate these models to the network conditions of the real world.

One apparent disadvantage of space searching procedures for a plane is that there is no way of making sure that facilities are sited in feasible locations, e.g. which avoid lakes or urban parks.

In practice, since the objective function for the p-median problem is shallow near the optimum as will be argued below, a nearby feasible location can be selected with little loss. Furthermore, these models are more likely to be used by planners to evaluate possible situations rather than to choose precise sites (Cargill and Hodgart, 1978). On a graph, feasible solutions can be defined as nodes so this difficulty does not arise.

In the empirical work on Edinburgh presented later, only the geometry of a plane is used. The reasons for avoiding a network geometry were mainly practical ones: most important, the task of handling the whole street network for a city as large as Edinburgh would be very expensive in both preparation and computer time. In defence of using a plane we can argue that unless there are major topographic barriers, Euclidean distance can be taken as an acceptable surrogate for the shortest path when users walk to a centre, provided the mesh of the street network is fairly fine in texture. Furthermore, for users who travel by car, there is very likely to be a roughly monotonic relationship between Euclidean distance and travel time, travel distance and even cost; there may even be a strong linear correlation. The existence of a monotonic relationship or a strong correlation would both provide support for taking Euclidean distance as a rough surrogate for distance in a network.

The most serious source of error with this surrogate is probably from journeys made by bus and train because the same distances between user and centre will have quite different times and costs when there is a convenient and frequent transport service on the route connecting them. However, to construct a data base which took full account of the whole transport network would be a substantial

study in itself. Since we are more concerned here with examining and developing models than with realistic applications, such an approach lies outside the scope of this thesis.

Another source of inaccuracy arises from the tapering fare structures common for public transport, whereby the second kilometer costs less than the first and so on. If the taper is consistent and smooth, it could be approximated by using an exponent of less than one for distance in computing cost of travel. Experiments carried out using such exponents, however, made little difference to the relative values of travel cost to a number of potential sites for new swimming pools, probably because the relation between travel cost and distance was still inherently monotonic. Moreover, many users walk to swimming pools (Currie, 1977) and are not affected by fare structures. For this reason, the non-linear aspect of fare structures was ignored in the work on pools.

6.1 Combinatorial programming

We can divide methods of solving the location problem into exact and heuristic approaches according to whether they result in a global optimum or an approximation to it. It is convenient to begin the discussion of exact methods by examining ways of solving the p -median problem on a graph and then discussing whether these methods can be used to solve problems with different goals. Suppose m facilities are to be located to serve n demand nodes on a graph. First of all we can ask how many different ways can the facilities be allocated. Since the centres should be sited at nodes by Hakimi's theorems, this question is the same as asking how many ways can m items be placed in n cells, a straightforward combinatorial problem to which the answer is

$$C_{n \ m} = \frac{n!}{(n-m)! \ m!}$$

In this formula $C_{n \ m}$ is inherently explosive in magnitude as n increases. For instance, there are over six and a half million distinct ways of assigning only 3 facilities to the 342 cells formed when Edinburgh is divided into 500 metre grid squares.

The combinatorial character of the problem derives from the fact that any node can be given the value '1' if it has a facility and '0' if it has not. A very clear exposition of methods of optimising a wide variety of spatial problems like this, which can be formulated as a structure of zero-one variables, is given in a series of studies by Scott (1969, 1971 and 1975).

In order to formulate the problem in combinatorial terms a binary variable, a_{ij} , is used to describe the way demand nodes are assigned to supply points. In the p -median problem all the demand from any given node must be assigned to the nearest facility, otherwise the solution is clearly sub-optimal. If j is the nearest facility to i , then all its demand assigns to j and a_{ij} must be one; if j is not the nearest facility then a_{ij} must be zero. Hence all assignments in the system and the associated catchment areas can be defined by this zero - one variable. On the assumption that there are no constraints on the capacity of facilities, the objective function for a ' p -median' (i.e. an m -median) problem becomes:

$$\text{minimise } Z_A = \sum_{i=1}^n \sum_{j=1}^n p_i d_{ij} a_{ij} \quad (5.1)$$

Here a_{ij} ensures in effect that travel from i to j is counted only when j is the nearest facility to i .

The main constraints in the problem are that only m facilities can be assigned and that all demand must be met. The whole set

of constraints, derived from ReVelle and Swain (1970), can be expressed:

$$\sum_{j=1}^n a_{ij} = 1 \quad (i=1, 2, \dots, n) \quad (5.2)$$

$$a_{jj} \geq a_{ij} \quad (i=1, 2, \dots, n) \quad (5.3)$$

$$(j=1, 2, \dots, n) \text{ and } i \neq j$$

$$\sum_{i=1}^n a_{ii} = m \quad (5.4)$$

$$a_{ij} \geq 0 \quad (i=1, 2, \dots, n) \quad (5.5)$$

$$a_{ij} = (0,1) \quad (j=1, 2, \dots, n) \quad (5.6)$$

The first constraint ensures that all demand at i is met by ensuring that one of the a_{ij} has the value 1. In models where allocation to supply points is probabilistic, a_{ij} would no longer be a 0 - 1 variable but would be defined instead as the fraction of i 's population assigned to j . The constraint would still hold.

The second constraint (5.3) ensures that the demand of a node with a facility will assign to itself rather than elsewhere, self assignment being denoted by a_{jj} or a_{ii} . When j has a facility a_{jj} must be 1, but a_{ij} may be 0 or 1. If there is no facility at j then both a_{jj} and a_{ij} will be 0. This constraint therefore prevents a_{jj} from having the value 0 when a_{ij} is 1. The third constraint (5.4) restricts the number of facilities in the system to m by making use of the fact that there must be m self-assigning nodes in the final solution.

Combinatorial programming involves the creation of a combinatorial tree on which each vertex describes one unique set of values for the variables to be solved, here the a_{ij} . A comprehensive examination of the tree is then carried out using a systematic searching procedure

such as the branch and bound or backtrack method. By this means the objective function is evaluated explicitly or implicitly at every vertex. Implicit evaluation means that it is inferred at a particular vertex that the optimal solution cannot lie on any branches descended from that point and so these branches are pruned out of the tree, thereby saving computation time. The most efficient searching strategy is therefore one where a high proportion of nodes are evaluated implicitly.

In reviewing these methods Scott (1971) argues that their flexibility enables them to be adapted to a great variety of problems in spatial planning. This attractive quality, however, may not extend to problems of locating facilities. It seems to be difficult in the latter problem to eliminate any feasible set of locations by implicit evaluation, though Ostresh (1973) presents several interesting attempts to use the spatial properties of the problem on a plane to do so. With the partial exception of Ostresh, no success in this direction has been recorded so far in the literature. Consequently, if little implicit enumeration can be done, tree searching virtually involves exhaustive evaluation of the huge number of possible solutions. Scott (1971) does not seem to give sufficient emphasis to this drawback of tree searching methods.

Although one can be sure of obtaining an optimal answer by exhaustive search, the cost in computation time for large problems can be prohibitive. Scott (1969) cites one case of dividing only twenty-five randomly-distributed demand points into two catchment areas which took fifty minutes of computer time using a backtrack algorithm. For this reason researchers are usually forced to employ

heuristic methods which obtain relatively good solutions by a rapid but incomplete search. However, an advantage of combinatorial methods is that they do permit the use of non-linear objective functions since they only require the objective function to increase or decrease monotonically as solution variables (i.e., the a_{ij}) are added. In theory they can therefore solve all the preceding formulations of the problem, including the demand-maximising problem, whereas linear programming can apparently be used only for the p-median and covering problems (Church and ReVelle, 1974). The advent of combinatorial programming and large high-speed computers has stimulated work on various spatial problems which were formerly daunting in size and complexity. Perhaps their most useful feature is that they provide optimal answers for comparison with the approximate solutions obtained by heuristic methods, thereby allowing the efficiency of the latter to be tested for small problems. As yet they provide little help with large problems.

Linear programming

A rather surprising feature of the p-median problem, in view of its integer character, is that it can be solved fairly easily as a linear programming problem. This has been demonstrated very neatly by ReVelle and Swain (1970). The constraint set for the preceding problem is actually from their formulation but they omit the last constraint which requires a_{ij} to be one or zero. Instead ReVelle and Swain take a_{ij} as the fraction of i 's population assigned to j ; this removes the integer element leaving all the constraints in a linear form. The problem therefore meets the basic requirements for solution by linear programming.

Although a_{ij} can now in theory have any value between 1 and

0 , in fact node i invariably assigns wholly to the nearest facility in order to minimise the objective function, thus yielding the desired solution of 1's and 0's . Hence the spatial structure of the problem forces a binary solution into the linear formulation.

The efficiency of linear programming essentially derives from the fact that a set of linear constraints defines a region in the n -dimensional space of the problem; all the feasible solutions to the problem lie inside that region. Since it is bounded by straight lines, planes or hyperplanes, this region must be convex in shape. The fundamental theorem of linear programming states that the optimal solution must be a corner point on this set so an algorithm for solving the problem need search only a relatively small number of points. If the constraints are treated as linear equalities, the solutions to this system of equations gives the desired corner points. The most widely used procedure in linear programming, the simplex method, can be regarded as a very efficient means of searching these corner points. This method affords a marked contrast to tree searching which usually requires a fairly large number of feasible solutions to be evaluated.

Since the 1940's linear programming has found a wide range of applications. One type of linear program, the transportation model, has been applied to several problems in geography and spatial planning (Garrison, 1959; Cox, 1965; Gould and Leinbach, 1966). In this problem goods have to be shipped from a set of origins to a set of destinations at minimal cost. Given the amount available at each origin, the amount needed at each destination and the cost of transport on all possible routes, the model yields the desired pattern of shipment. In geography this model has been used to find the optimal

catchments around central facilities of known capacities whose locations are fixed (Yeates, 1963). Because the locations must be fixed, the transportation model cannot be used directly to solve the location problem, but it can be used to explore the effect of constraints on capacity (Gould and Leinbach, 1966).

The ingenuity in ReVelle and Swain's work lies in adapting the simplex model to solve the location problem, in spite of the latter's integer character and the apparent lack of a ready means for incorporating the location variable into the structure of the model. Detailed expositions of linear programming, including the simplex and transportation models can be found in numerous texts (Chung, 1963; Loomba, 1964) including a basic text on geography (Abler et al., 1971).

ReVelle and Swain were able to solve a small test problem with 30 communities and 6 centres in only 1.51 minutes of computer time. Their solution only needed 173 iterations, a very small fraction of the 593,000 possible solutions. The efficiency of the method for large problems has not been reported but it will clearly be far faster than tree searching.

The dual variable of a linear program can provide useful information on the marginal benefit which would accrue if extra units of resources were made available. In this case the dual variable for the constraint on the number of centres would indicate how average travel distance would change, given one more centre. The problem of locating additional facilities can also be solved quite easily by this method. The presence of an existing facility is indicated simply by making the a_{ii} for that node 1.

Extensions of the linear programming formulation

In an extension of this model Rojeski and ReVelle (1970)

replace the constraint on the number of centres by a constraint on the funds available to open new facilities or expand those already opened during the process of solution. If costs of opening and costs of expansion vary among the potential locations, this constraint has the form:

$$\sum_{j=1}^n f_j a_{jj} + \sum_{j=1}^n b_j \sum_{i=1}^n p_i a_{ij} \leq C$$

where f_j is the fixed cost of opening facility j

b_j is the variable cost of expanding j by one
unit of population or demand

C is the investment budget

The first term gives the cost of opening or constructing new facilities. In the second term $\sum_i p_i a_{ij}$ is the actual population assigned from all other nodes to facility j i.e. the capacity of j . The associated expansion cost, a kind of operating cost for meeting this level of demand, is $b_j \sum_i p_i a_{ij}$, so the second term gives the running costs for that set of facility locations.

This extended model has the advantage of allowing data on the costs of supply to be incorporated, if available. Although the inclusion of an implicit constraint on capacity seems appropriate for many services (e.g., hospitals and schools), it means that some customers may not be allocated to their most convenient facility but rather to the cheapest one (in terms of supply costs) which is not already fully used. It is very unlikely that a predetermined budget will fortuitously allow an exact whole number of facilities. Let us suppose the optimal solution is 9.5. To remove this difficulty Rojeski and ReVelle provide the planner with a means of finding out what budgets correspond to 9 and 10 facilities. These alternative

integer solutions can then be compared in terms of efficiency. In fact, by solving the problem for a range of budgets, the authors argued that they were able to provide the potential decision maker with information on the trade-offs between funds allocated and the travel costs of users in the system. The shape of this trade-off curve resembled the profile of the diagram relating m and aggregate travel, discussed earlier (Fig. 3.4), though the former was concerned with a network and the latter with a plane.

To locate additional facilities, the a_{ii} corresponding to those facilities already in existence are set to one, as before. For this problem and the general problem Rojeski and ReVelle's model is obviously more flexible and satisfactory analytically than the previous model because it takes capacity constraints into account and allows the option of expanding existing facilities as well as opening new ones. However, its data requirements are formidable. These could be met only by an intensive study or by data from the accounts of a cost-conscious service, but this largely precludes its use as a general instrument for exploring provision of services in an area, unlike the preceding model. Applications to the real world have not yet been reported. Finally, it is worth noting that linear programming is unlikely to be suitable for use-maximising models because the objective function, based on a term like $\sum p_i e^{-bd_{ij}}$, is non-linear.

Models for maximising social benefit

Wagner and Falkson (1975) criticise the preceding models for assuming that demand is inelastic and argue that public services with inelastic demand are the exception rather than the rule. They then present a series of elegant models which balance the net benefits to

consumers of receiving the service against the marginal cost of supplying it, thereby maximising both consumers' and producers' surpluses. Though interesting conceptually, it is quite hard to envisage these models being made operational; the authors in fact were only concerned with formulating the models and did not present any applications. In a sense the latter parts of Chapter 3 and of Chapter 4 make part of their model operational, albeit for an isotropic plane.

Because the elasticity of demand has been ignored, Wagner and Falkson argue that decisions based on fixed requirement models will continually result in an excessive amount of the service being supplied. In view of its significance for public policy, it is worth drawing attention to several flaws in this argument, though it should be noted that the authors do not make their grounds completely explicit. First, the fixed demand models just discussed do not determine the level of supply; rather they assume that the amount to be supplied, usually a number of facilities, has previously been determined in the political forum or elsewhere. It is hard to see why the supply determined in this way should necessarily be more than would result from Wagner and Falkson's model; it could be less.

Second, several essential public services do have inelastic demand - schools and fire protection, for instance. Third, on a philosophical level, to rely only on the Pareto optimum embedded in their model is to overlook the criteria of need and equity which are both essential to the allocation of all basic public services (Olsson, 1974; Dear, 1974). Even where demand is elastic, a value judgement could be made legitimately to prefer a fixed demand model because it selects more equitable locations.

On a more technical point, Wagner and Falkson appear to infer from evidence of distance decay that demand is elastic. As noted previously, distance decay rings around centres may not be inconsistent with inelastic demand, because they may only reflect declining preference for that particular centre. Judging by this, services with elastic demand will be less common than Wagner and Falkson appear to think.

In spite of these criticisms Wagner and Falkson's argument does agree with the main conclusion reached in Chapter 4 about the influence of elastic demand on the optimum value of m under the most likely conditions of supply (Example C). It was also noted, however, that there can be exceptions to this general conclusion, depending on supply conditions and on the way supply is treated in the model. Also, in certain cases (Example E) cost minimising models could underestimate the advantages of organising a given supply in smaller, more dispersed units.

To be fair, this contention of Wagner and Falkson is actually somewhat peripheral to the main matter in their paper which is the presentation of an elegant and interesting model, even more general than Rojeski and ReVelle's, but even harder to apply. Nevertheless, their criticism of the purely locational models just discussed is not, strictly speaking, valid.

Conclusion

In many respects, standard methods of optimisation such as linear and integer programming have been very helpful as a means of formulating the location problem in a form which makes it easier to examine and solve. To judge by the existing literature, however, they are not generally practicable for large problems in a network, because

they require too much time in computation. Fortunately there are other efficient methods available for solving problems on a plane; these will be examined in the next chapter.

CHAPTER 6

Methods of searching space on a plane

Introduction

Whereas in a network feasible sites are restricted to a finite number of nodes, on a plane the number of potential locations is infinite, by definition. As noted earlier, methods of locating centres in a network could be applied to a plane, but they will usually be inefficient, partly because the number of possible locations is so great and partly because they do not really exploit the properties of a plane. In this chapter we will discuss how to take greater advantage of the spatial information inherent in the geometry of a plane.

Consider a case where one centre is to be located on a plane. No matter which of the objective functions is being used, we can evaluate the function at any point on the plane surface. Contours joining points with the same value can then be drawn and we can therefore think of the surface as having a gradient. At each location that gradient will point towards areas where the objective function is higher or lower. The slope can therefore be used continuously as a guide in searching for peaks or hollows, depending on whether the problem is to find the maximum or minimum. This property, of allowing the gradient to guide search, is in fact the basis of algorithms by Cooper (1963) and Törnqvist (1971) for solving the p -median problem on a plane.

Surface topography of different objectives

Almost all the published work on methods of searching space is concerned with the p -median problem. Before considering how to use such methods for other objectives we need to consider the nature of the

surfaces which the other objectives produce. For this purpose it is convenient to use a very simple example, taken from Hodgart (1978).

The distribution of population by villages along a narrow isolated mountain valley is shown in Figure 6.1. We wish to determine the best locations for a number of services which have not hitherto been provided in the valley; one facility is to be constructed for each. All movement to facilities is to be on foot; travel cost is therefore a linear function of Euclidean distance. To simplify discussion, we restrict potential sites to certain points on the plane, the villages themselves, but this does not essentially affect our conclusions.

The value of aggregate travel for each village, computed for a service with inelastic demand, is shown in Figure 6.2. With this objective the surface of the profile is smooth and runs continuously down towards the single minimum at E. If we move down the slope from any starting point until the gradient is 0 or turns positive, it seems that we can be sure of reaching the global minimum.

Aggregate travel surfaces have in fact been computed for many areas (Harris, 1954; Neft, 1966). Invariably they have one optimum and slope smoothly towards it, so that a marble placed on the surface would inevitably roll down to the minimum. In Figure 6.2 the gradient is gentle near the minimum, so that little efficiency would be sacrificed by locating at D or F or even G. Similarly, Cooper (1963, p. 340), Eilon et al. (1971), Nordbeck and Rystedt (1972) and, in slightly different context, Goodchild (1972) make the important point that the aggregate cost surface is shallow in a certain region around the minimum for very different distributions of population.

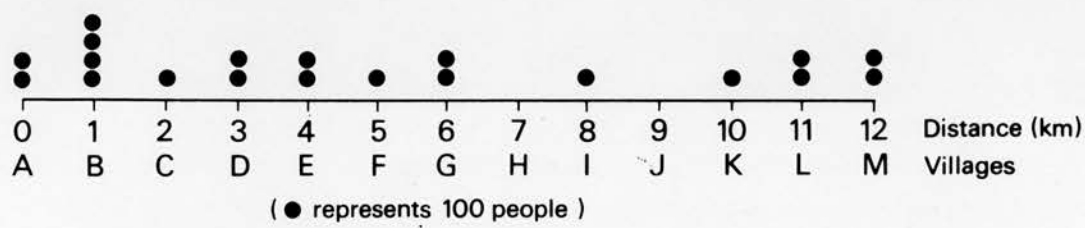


Figure 6.1 Distribution of population in hypothetical example

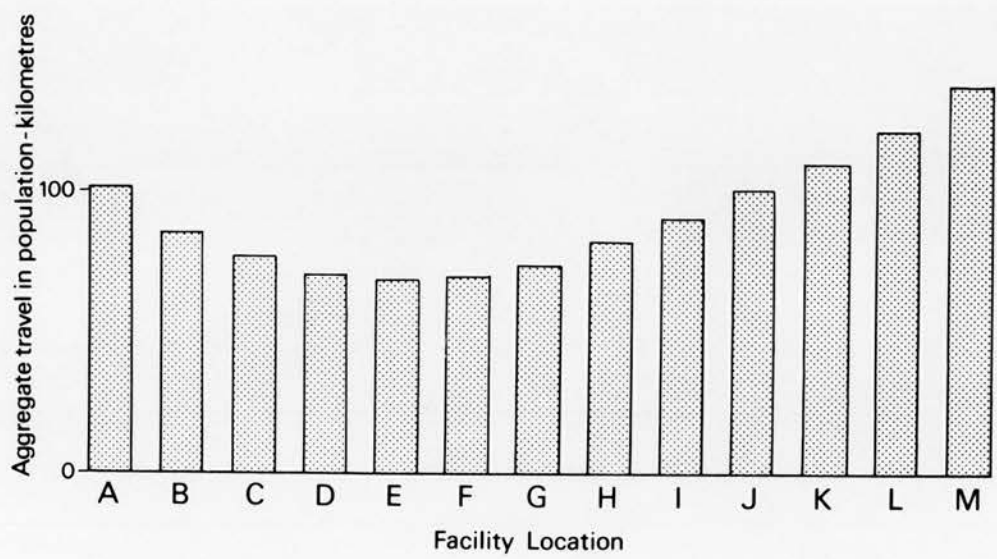


Figure 6.2 Aggregate travel at each village

To minimise mean travel we minimise

$$Z_B = \frac{\sum_{i=1}^n p_i d_{ij}}{\sum_{i=1}^n p_i}$$

Now, though d_{ij} varies with the location of a centre, $\sum_{i=1}^n p_i$ is a constant for all locations and Z_B therefore has only variable term namely $\sum_{i=1}^n p_i d_{ij}$. Hence, the location which minimises aggregate travel also minimises mean travel. The topography of a surface for mean travel would therefore be identical to that for aggregate travel in relative terms.

The essential properties of an aggregate travel surface are not affected by disaggregation. Suppose, for instance, that the population in each village is divided into ℓ groups. Let k denote a particular group; then $k=1, 2, \dots, \ell$. Let the average demand from any group, p_{ik} , during a certain time period be w_k . Demand from any single village is then

$$\sum_{k=1}^{\ell} w_k p_{ik}$$

so the objective becomes

minimise
$$Z_c = \sum_{i=1}^n d_{ij} \sum_{k=1}^{\ell} w_k p_{ik}.$$

It seems very reasonable to expect that the topography of this function's surface will also be smooth and have one centre.

If the objective is to maximise use of the facility and use depends on access, declining as e^{-bd} , we can compute an index of potential use at each village. When $b=1$ the profile is fairly uneven (Fig.6.3): there is a pronounced global maximum at B, a clear local maximum at L and weakly defined maxima at D, G and I. Because remote population is now being discounted, the profile is very sensitive to local pockets of population as is the case with population potential surfaces (Harris, 1964; Neft, 1966; Nordbeck and Rystedt, 1972). Therefore a hill-climbing method of search which

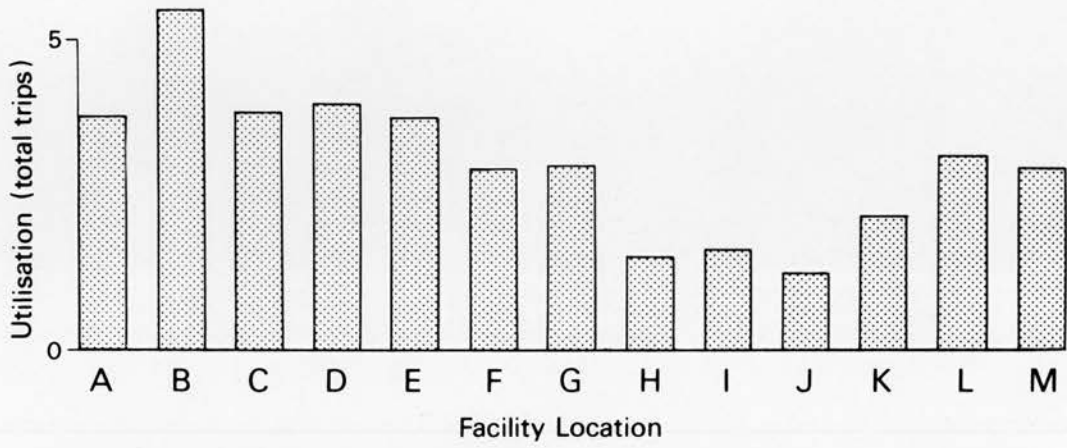


Figure 6.3

Level of use of a facility sited at each village
when $b = 1.0$

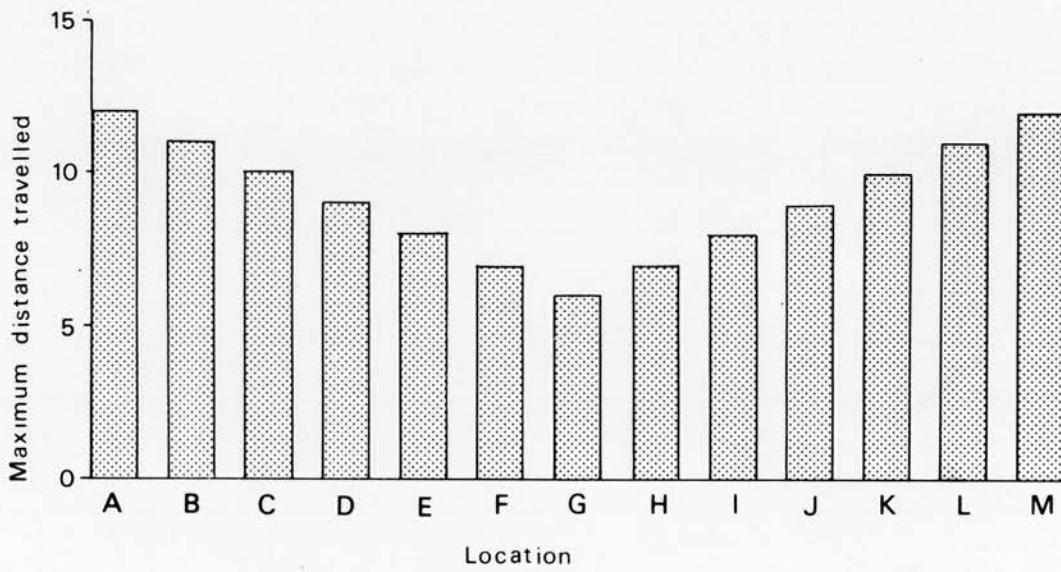


Figure 6.4

Maximum distance travelled to a facility sited at
each village

moved in the direction of ascending gradient might find only a local summit, depending on where it started. The chance of finding the global optimum could then be improved by using a series of different starting positions.

The degree of spatial inequity at each potential site can be measured by taking the difference in travel distance between the nearest and most distant members of the population. As Figure 6.4 confirms, the surface for this goal has intrinsically linear slopes, falling towards a single optimum where inequity is minimised.

The centre of gravity of the area's population can be calculated directly as the mean of the weighted co-ordinates (Fig. 6.5). Since it minimises the sum of the squared travel distances, it gives more weight to extreme distances and therefore produced a somewhat more equitable solution than the median. Since the mean can be found without a searching procedure the topography of this surface, based on d_{ij}^2 , is not of interest in the present context.

A covering objective can allow equity to be treated more flexibly than the minimax criterion; it would be particularly relevant to certain services such as fire protection. To evaluate this objective at any village, we ascertain how many people (or households) are within the specified covering radius, S . As Figure 6.6 shows, when $S = 3$ km the surface has a global maximum at D and a local maximum at J . In general, when S is small, distant population is discounted and an uneven topography resembling Figure 6.3 is produced. When S is large the surface will be smoother with fewer peaks, tending to resemble Figures 6.2 or 6.4.

To summarise, for certain objectives a space-searching strategy guided by the gradient may find a purely local optimum. The more elastic the demand the more numerous the local optima on the demand

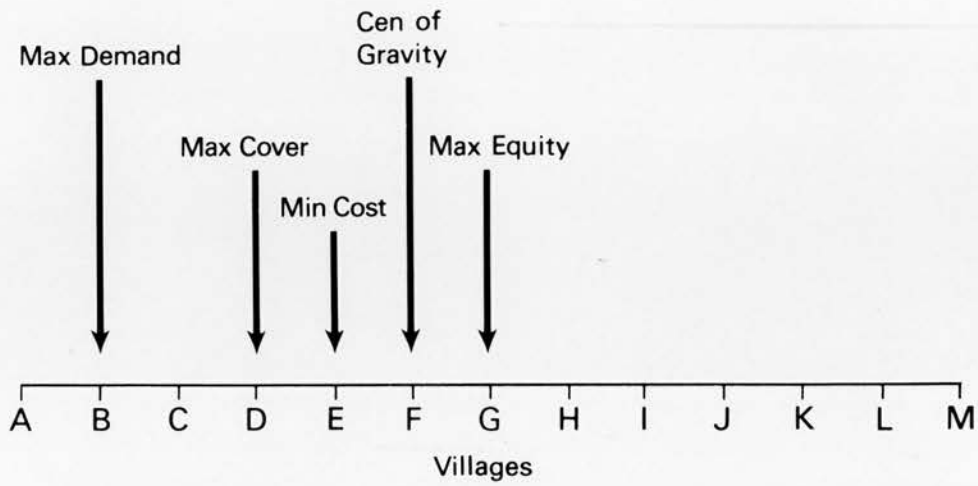


Figure 6.5 Optimal locations derived from various objectives

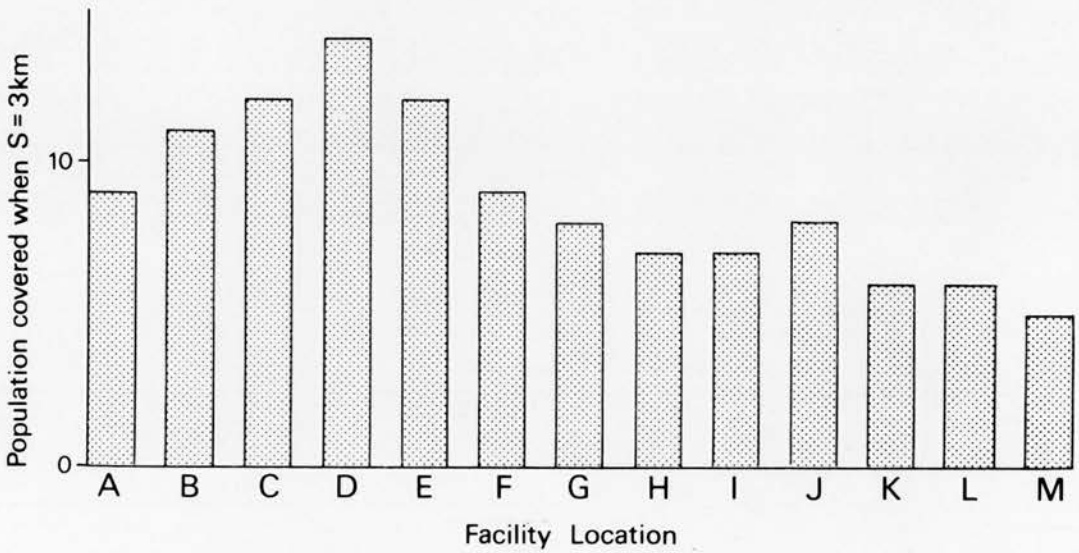


Figure 6.6 Population covered from a facility sited at each village when $s = 3$

surface. Likewise, for a problem involving maximal cover, the smaller the value of S , the greater the danger of finding a local optimum. Though we have demonstrated these properties with a one dimensional distribution for ease of discussion and illustration, our conclusions will obviously apply also to a plane. Clearly, the p-median problem is much easier to solve by space searching than the use-maximising or covering problems because its topography is much simpler.

Locating one centre on a plane

The core of many algorithms for minimising travel on a plane is a rapid method for finding one centre, devised independently by Weiszfeld (1937), Miehe (1958), Cooper (1963) and Kuhn and Kuenne (1962). It was suggested earlier that the gradient at any point would indicate the direction of the optimum. Essentially, the method developed by these writers is based on finding a mathematical expression for the gradient. Because of its general interest both as a method of searching space and of solving other formulations of the problem, an account of this method is given here.

We consider how the aggregate distance travelled (AT_i) to a centre (X_j, Y_j) by the population, p_i , of any point i with co-ordinates (x_i, y_i) changes as the centre's location moves in Euclidean space.

By definition

$$AT_i = p_i d_i = p_i [(x_i - X_j)^2 + (y_i - Y_j)^2]^{\frac{1}{2}} \quad (6.1)$$

As the centre moves, x_i, y_i and p_i are fixed and can therefore be treated as constants. Hence AT_i can be regarded as a function of the variables X_j and Y_j i.e. $AT_i = f(X_j, Y_j)$.

The gradient of this function can be described by the partial derivatives with respect to X_j and Y_j . Since the partial derivative with respect to X_j describes how AT_i or $f(X_j, Y_j)$ responds to an infinitely small increase in X_j when Y_j remains constant, to obtain this derivative we can treat AT_i as $f(X_j)$. The partial derivative we wish to obtain is then

$$\frac{dAT_i}{dX_j} = \frac{df(X_j)}{dX_j}$$

To obtain the gradient we first substitute

$$u = d_i^2 = x_i^2 - 2x_i X_j + X_j^2 + y_i^2 - 2y_i Y_j + Y_j^2 \quad (6.2)$$

into 6.1. AT_i , as $p_i d_i$, now becomes $p_i u^{\frac{1}{2}}$. Thus AT_i is a function of u .

We can define this function as $g(u)$. By definition u is a function of X_j (6.2); we can call this function h , so u can be treated as $h(X_j)$. Now since $AT_i = g(u)$ and $u = h(X_j)$ aggregate travel is a function of a function of X_j .

Using the chain rule for differentiating a function of a function (Wilson and Kirby, 1975, p. 136)

$$\frac{dAT_i}{dX_j} = \frac{dAT_i}{du} \cdot \frac{du}{dX_j} \quad (6.3)$$

Of course u was chosen so that both derivatives on the right hand side of (6.3) would be straightforward:

$$\frac{dAT_i}{du} = \frac{1}{2} p_i u^{-\frac{1}{2}} = \frac{p_i}{2[(x_i - X_j)^2 + (y_i - Y_j)^2]^{\frac{1}{2}}} = \frac{p_i}{2d_i} ;$$

$$\text{and } \frac{du}{dX_j} = -2x_i + 2X_j = -2(x_i - X_j) .$$

Hence

$$\frac{dAT_i}{dX_j} = \frac{p_i (-2)(x_i - X_j)}{2d_i} = -\frac{p_i (x_i - X_j)}{d_i} \quad (6.4)$$

Applying the same method, the partial derivative of AT_i with respect to Y_j is

$$\frac{dAT_i}{dY_j} = \frac{-p_i (y_i - Y_j)}{d_i}$$

If Z_A is the aggregate travel from all n points to the centre then

$$Z_A = \sum_{i=1}^n AT_i$$

From (6.4) an infinitely small increase in X_j will cause each point to contribute a certain increment in travel, given by the partial derivative at that point. The overall rate of change (i.e. the gradient of Z_A) will be the sum of these contributions

$$\text{i.e.} \quad \frac{dZ_A}{dX_j} = \sum_{i=1}^n \frac{-p_i (x_i - X_j)}{d_i}$$

At the minimum point this gradient is zero by definition:

$$\sum_{i=1}^n \frac{-p_i (x_i - X_j)}{d_i} = 0 \quad (6.5)$$

By manipulating (6.5) it can be shown that

$$X_j \sum_{i=1}^n \frac{p_i}{d_i} = \sum_{i=1}^n \frac{p_i x_i}{d_i}$$

Hence

$$X_j = \frac{\sum_{i=1}^n \frac{p_i x_i}{d_i}}{\sum_{i=1}^n \frac{p_i}{d_i}} \quad (6.6)$$

By symmetry

$$Y_j = \frac{\sum_{i=1}^n \frac{p_i y_i}{d_i}}{\sum_{i=1}^n \frac{p_i}{d_i}} \quad (6.7)$$

These expressions for the location of the minimum point cannot be solved directly because d_i is unknown. Kuhn and Kuenne found in practice, however, that it could be solved fairly easily by iterative

approximation. The latter method consists of substituting arbitrary initial values of X_j and Y_j which allow the right hand side of 6.6 and 6.7 to be evaluated. This yields new values of X_j and Y_j which, as facility co-ordinates, have lower aggregate travel in fact than the initial values and can in turn be re-substituted back into 6.6 and 6.7 to yield a further improvement. When no further improvement is obtained the iterative procedure has converged on the minimum.

With an iterative solution there is always the possible danger that the procedure may not converge. For this procedure, however, cases of non-convergence have never been reported in the literature. In practice convergence is usually quite rapid, especially if a good initial approximation such as the centre of gravity is used. A second danger is that the minimum will be local rather than global. Since we have demonstrated, albeit graphically, that aggregate travel surfaces for one centre are smoothly concave away from a single minimum, by descending the slope from any starting point the optimum will be reached almost inevitably. A thorough analysis of these and other mathematical properties and problems of the method is given by Ostresh (1973).

In contrast, the method developed by Törnqvist (1971) obtains a measure of the gradient by calculating aggregate travel to an initial trial cell and then to another cell one grid position to the west. Movement along an east-west axis then along a north-south axis continues until no improvement is recorded. The search can be carried out in successively smaller steps for greater accuracy. A complete account of this method is presented by Kohler in Rushton et al (1973). If the surface of the objective function is smoothly concave away from

the optimum this method will be just as successful in finding the minimum as the previous one, though it needs slightly more computation time to detect the direction of the minimum.

Exaggerating a little, we can liken the two methods to bears searching for a barrel of honey on a pitch black night. The iterative method is like a bear with a quick and accurately directed sense of smell. On the other hand the Törnqvist method is like a bear which has a bad cold and a much more weakly directed sense of smell and therefore has to take several steps in each direction at any stage to check whether it is getting closer to the honey. Clearly the iterative bear will usually find the honey more quickly.

This general strategy of search outlined above could be used to optimise other objective functions. For instance, if the number of trips to a facility declines exponentially with distance so that T_i is given by $p_i f_0 e^{-bd_i}$ (after 3.10.1) then it can be shown (Appendix 2) that the partial derivatives for X_j and Y_j would be:

$$\frac{dT_i}{dX_j} = bp_i (x_i - X_j) e^{-bd_i} d_i^{-1} \quad ; \quad (6.8)$$

and

$$\frac{dT_i}{dY_j} = bp_i (y_i - Y_j) e^{-bd_i} d_i^{-1} \quad (6.9)$$

From this it can be shown that at the point where trips are maximised

$$X_j = \frac{\sum_{i=1}^n p_i x_i e^{-bd_i} d_i^{-1}}{\sum_{i=1}^n p_i e^{-bd_i} d_i^{-1}} \quad (6.10)$$

and

$$Y_j = \frac{\sum_{i=1}^n p_i y_i e^{-bd_i} d_i^{-1}}{\sum_{i=1}^n p_i e^{-bd_i} d_i^{-1}} \quad (6.11)$$

In general the greater the absolute value of b , the more numerous become the local peaks on the surface of $\sum_i T_i$ and consequently the

more likely is any searching procedure to stick on a local optimum. Computational experience with iterative solution of equations like 6.10 and 6.11 has not so far been reported in the literature. Though some points where the partial derivatives are zero could in theory be minima or saddle points, in the computational experience which will be reported in the next chapter only maxima were found.

Similarly, if the net benefit to users at i of a facility at j is

$$\begin{aligned} \text{NBU}_i &= vT_i - T_i d_{ij} \cdot c_t \\ &= v f_{oi} p_i e^{-bd_{ij}} - f_{oi} p_i e^{-bd_{ij}} \cdot d_{ij} c_t \end{aligned}$$

then it can be shown (Appendix 3) that

$$\frac{d\text{NBU}}{dX_j} = p_i e^{-bd_{ij}} (x_i - X_j) (vbd_{ij}^{-1} + d_{ij}^{-1} - b)$$

The co-ordinates of the point where NBU is maximised are then given by

$$X_j = \frac{\sum x_i p_i e^{-bd_{ij}} (vbd_{ij}^{-1} + d_{ij}^{-1} - b)}{\sum p_i e^{-bd_{ij}} (vbd_{ij}^{-1} + d_{ij}^{-1} - b)} \quad (6.12)$$

and

$$Y_j = \frac{\sum y_i p_i e^{-bd_{ij}} (vbd_{ij}^{-1} + d_{ij}^{-1} - b)}{\sum p_i e^{-bd_{ij}} (vbd_{ij}^{-1} + d_{ij}^{-1} - b)} \quad (6.13)$$

As with 6.10 and 6.11, the larger the value of b , the more numerous the local peaks on the surface of $\sum_i \text{NBU}_i$ and the greater the probability of finding a local optimum from iterative solution. As before, only maxima were found in computing the iterative solutions.

Space-searching heuristics for m facilities.

The most efficient heuristic for solving p -median problems on a plane appears to be the 'alternate' algorithm developed by Cooper

(1963, 1967, 1968), so called because it alternates between allocating population and locating centres in the following sequence:

- (1) assign each centre to an arbitrary initial location;
- (2) allocate each demand point to its nearest centre, thereby defining m Thiessen polygons (allocation stage);
- (3) relocate each centre to the median within its catchment area by the Kuhn-Kuenne method (location stage);
- (4) repeat steps (2) and (3) until convergence.

Convergence must occur because steps (2) and (3) have the same objective; each can only reduce aggregate travel. Cooper also referred to this as the 'location/allocation' method; apparently the term was then used by later authors to describe the class of problems itself.

With one centre the Kuhn-Kuenne algorithm can be reasonably sure of finding the global optimum, but with several centres the combinatorial nature of the problem creates a much vaster tree to search. Local optima are therefore more frequent; i.e., the starting points may restrict the pattern of search, thereby overlooking the global optimum. This danger can be reduced by using a number of different starting points. Accordingly, Cooper (1964) tested his method quite rigorously, using one hundred trial problems each of which had forty randomly generated demand points and three centres. As an index of accuracy he took the difference in aggregate travel between the global optimum, O_i , and the approximate solution, H_i , for each problem, expressed at a ratio of the former. From this he obtained the following measure:

$$MPE = \frac{100}{n} \sum_{i=1}^n \frac{H_i - O_i}{O_i}$$

where MPE is mean percentage error and n is the number of trials.

Obviously, the closer MPE is to zero, the more accurate the heuristic. For the hundred trial problems MPE was only 2.582%.

If the algorithm was to be applied to the problem of locating three libraries in an area where none existed it is unlikely that the degree of accuracy required by the relevant planning body would be greater than 2.5%, because the algorithm would be used as a general guide to where the libraries should be situated not to their exact sites. The actual choice of sites would probably depend on availability of land, bus routes, parking space, planning constraints and so on. In practice it would seem best therefore to use the algorithm to generate a number of good approximate solutions and then assess such solutions according to how suitable nearby sites might be. An MPE of 2.5% is likely to be more than adequate for such an application.

Cooper has tested several refinements to the alternate method. One of these is designed to obtain a set of good starting points. Another takes the solution produced by the alternate method and "jumps" out of it into two other local optima. Neither appears to have been adopted by other researchers, perhaps because computer time is now less at a premium and the alternate method is quite efficient on its own. An 'alternate' type of algorithm for solving the p-median problem on a graph is given by Maranzana (1964). A somewhat different approach by Teitz and Bart (1968) solves the same problem by systematic substitution of new supply nodes for those already in solution at any stage. The latter method finds the optimum more frequently but requires more time for computation (Rushton et al., 1973).

The Törnqvist algorithm

The searching strategy of the Törnqvist algorithm has already been

described for the special case involving only one centre. The searching strategy for m centres, however, is harder to summarise clearly and adequately. Before the algorithm starts, maximum and minimum step sizes have to be specified. At any stage the search starts with the maximum step, changes to smaller steps for finer search and stops when the minimum step size has been reached. The algorithm can be outlined as follows.

1. Fix the location of all centres except one, denoted by F_1 . Determine all catchment areas. Evaluate Z and call this Z_1 .
2. The moveable centre, F_1 , takes one step (of the specified initial size) west. This corresponds to the location stage of Cooper's algorithm, except that it only applies to one facility.
3. New catchment areas are defined for all m centres and Z is re-evaluated. Call the new value Z_2 . This step is equivalent to the allocation stage of Cooper's algorithm.
4. If $Z_2 < Z_1$, replaces Z_1 by Z_2 . Go back to step 2 and repeat steps 2 to 4 until no improvement is recorded i.e. $Z_2 > Z_1$. When $Z_2 > Z_1$, proceed to step 5.
5. Move F_1 one step east from the best location found so far.
6. Define new catchment areas, evaluate Z and call this Z_2 .
7. If $Z_2 < Z_1$, replace Z_1 by Z_2 . Go back to step 5, repeating steps 5 to 7 until $Z_2 > Z_1$. Thereupon go to 8.
8. Halve the step size and go back to 2. Continue cycling back to step 2 until the step size is less than the minimum specified, at which point terminate the east-west

search and fix the easting (i.e., the y coordinate) of F_1 at the best value found.

9. F_1 now searches south and then north in the same manner as it searched east and west over steps 2 to 8. When $Z_2 > Z_1$ F_1 has found its best location, given the fixed locations of other centres.
10. Repeat steps 2 to 8 for each other facility in turn. When the search for the last facility has finished, one whole iteration of the algorithm has been completed.
11. Begin a new iteration by moving F_1 again as in step 2 and continue until no changes are made between successive iterations.

The description of the algorithm given by Törnqvist et al. (1971) is generally clear but is not sufficiently explicit for the detailed history of search printed out during the execution of the algorithm to be completely intelligible. The description given above, based on that given by Kohler (Rushton et al., 1973, is more complete and enables the searching sequence to be understood. Like the Cooper algorithm, NORLOC alternates between relocating the facilities and allocating the demand points. The allocation stage is virtually identical in both but, whereas Cooper relocates all facilities at once, NORLOC gingerly moves one centre one step, allocates, moves it another step then allocates again.

Comparison of Törnqvist and Cooper algorithms.

Since a comparison of the efficiency of the two algorithms has not so far appeared in the literature apart from a brief note in Hodgart (1978), a short assessment may be useful. The crucial

difference between the alternate procedure and the NORLOC algorithm is that NORLOC only relocates one centre at a time in its location stage, invoking the whole allocation stage each time a single centre moves one step. Since the allocation stage involves determining all the catchment areas afresh, this is expensive in computer time. When m is large the algorithm will in fact repeatedly check and recompute the same sums of population and distance, except for the few cells whose assignment has shifted. In contrast, Cooper's algorithm moves all the centres at the location stage and therefore obtains a more decisive improvement before invoking the allocation stage and evaluating Z . Moreover, by treating each catchment separately, Cooper's algorithm only has to compute distances within that catchment during the location stage.

Some smaller disadvantages of NORLOC may be noted briefly. At the location stage it can only change one co-ordinate at a time whereas the Kuhn-Kuenne method enables the x and y co-ordinates to be altered simultaneously from a given set of travel distances. Also, examination of NORLOC's searching sequence shows that it sometimes repeats earlier evaluations. For instance, it may move several steps west, stop, then lacking a knowledge of the gradient, it has to move back one step east and re-evaluate one of the locations just tested.

A possible advantage of NORLOC is that it does not require the objective function to have a derivative. Törnqvist's way of searching space might conceivably be used, therefore, with a wider range of objective functions.

From the above discussion it is clear that NORLOC will be more expensive in computer time. Nevertheless, there is no reason why it

should be less accurate in finding the optimum. Moreover, the previous criticisms should not be allowed to detract from Törnqvist's achievement as a geographer in devising a very effective means of solving a spatial problem which had evaded solution for some time in operations research and applied mathematics.

A small part of the empirical work presented below was carried out using NORLOC, despite its disadvantages. There were several practical reasons for this. The most important was that it was easier to make the program operational at Edinburgh. Several difficulties were encountered in compiling and running the LAP algorithm, obtained from Iowa, which in fact had to be modified slightly. At this stage in the work NORLOC was thought to be more flexible in tackling the additional facility problem. Later it became clear that this was wrong in principle, though true in practice because further difficulties were encountered with LAP on this particular type of problem.

It is surprising that a comparison of the two algorithms has not been published. Part of the reason probably lies in the fact that Cooper's method was better known in N. America particularly to workers in operations research and regional science, whereas Törnqvist's algorithm has been more popular with geographers in Scandinavia and, to a lesser extent, in Britain. In fact Törnqvist's work was originally published in Swedish in 1964 and remained unknown outside Scandinavia until Gould's discussion of it (Abler et al., 1971; Törnqvist et al., 1971).

Since the additional facility problem is more common in reality than the general problem it is important to assess how accurately any heuristic will solve this type of problem. Both LAP and

NORLOC tackle this problem by tying existing facilities to their locations and allowing the search for additional ones to begin from their starting positions. Local optima now become more frequent because the fixed locations of existing centres constrain the search. In Figure 6.8 the new facility will be unable to reach the global minimum at O from a starting position at S because if it moved towards O aggregate travel would increase as it encroached on the catchments of the existing centres, fixed at A and B . Use of a Cooper or Törnqvist type of procedure with other objective functions will encounter the same difficulty. Repeated trials with various carefully chosen starting locations will probably be needed then to identify the optimum.

Experiment with searching of interstices

When the existing number of facilities is fairly large it may be cumbersome to experiment with numerous sets of starting locations. In such circumstances the most promising locations for new centres would seem to be the interstices between existing centres. It was therefore felt that an algorithm which systematically evaluated all these interstices might be a more convenient way of solving this problem than LAP or NORLOC.

Accordingly an attempt was made to write such an algorithm. However, the concept of an interstice proved very difficult to define in terms which could be made operational on a computer. Given a set of points (Fig. 6.9a) one can immediately draw by eye a framework of triangles joining them and hence take the centre of the triangles as the interstices (Fig. 6.9). However, the failure of repeated attempts suggests that in this case the eye's intuitive skills in recognising geometrical patterns are very hard to translate into the algebraic logic of a computer language. This difficulty was

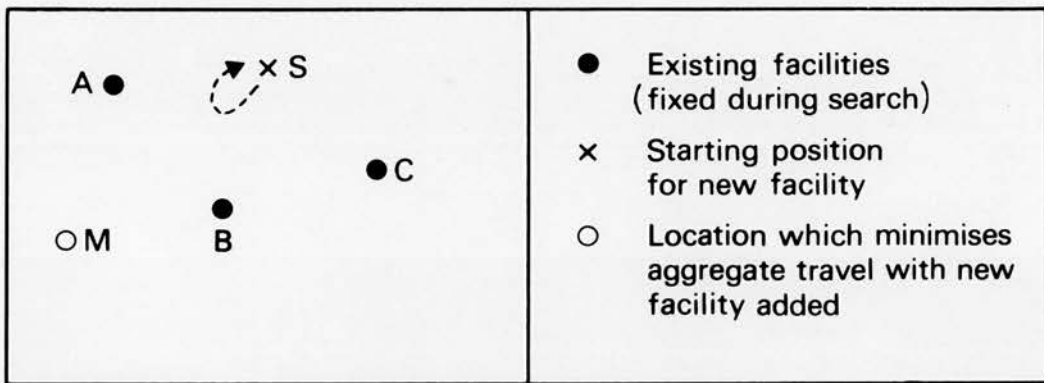


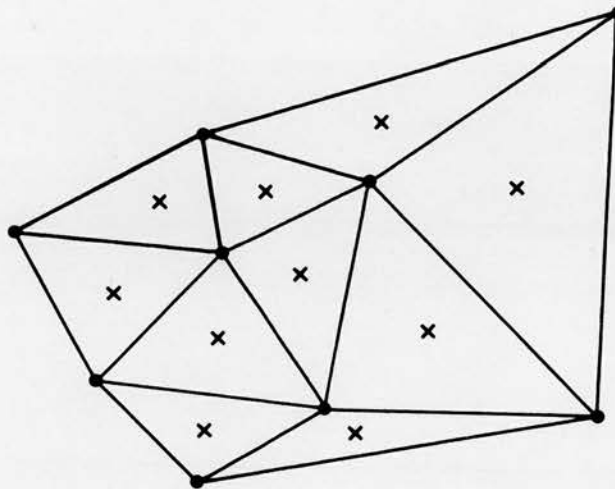
Figure 6.7

Search restricted by location of existing facilities

(a)



(b)



x : interstices

Figure 6.8

(a) Locations of a given set of points

(b) Construction of a framework of triangles and definition of interstices

eventually bypassed by reading the triangular framework in as data, a cumbersome solution.

It might be possible to combine the skills of the eye and the numerical power of the computer through interactive computer graphics. Given a visual display of the distribution of demand and the location of existing facilities, the researcher or planner could pick promising locations by eye and have them evaluated fairly rapidly in terms of various indices of efficiency and equity. Interactive graphics would seem to offer a promising avenue for future work.

A more mathematical means of solving this problem might be through the use of Delaunay triangles. Delaunay triangles are the geometrical duals of Thiessen polygons: they are formed by drawing straight lines between points whose Thiessen polygons have one side in common (Boots, 1974). An algorithm for defining the Delaunay triangles for a given set of points has been developed by Boots (1974). Use of such an algorithm would probably solve the problem of defining the initial triangular framework.

Though the attempt to develop an interstice searching algorithm failed to produce a method which was sufficiently practical to help in solving the locational problem, from this exercise it was realised that the rationale underlying existing algorithms needed closer examination and criticism. The exercise therefore led indirectly to the development of the models elaborated in Chapters 3 and 4.

In conclusion, the alternate method developed by Cooper seems to provide the best general approach to the problem of locating facilities on a plane. Hitherto, this method has only been used with the objective of minimising travel cost within Thiessen polygons. It is not obvious how well the method might work with the goals of maximising use or

net benefit to users. Having derived mathematical methods for optimising both of these goals, it therefore seemed worthwhile to test these within the framework of an alternating algorithm. It also appears that no one has tried to use Cooper's method with overlapping catchment areas. Accordingly, the next chapter is devoted to attempts to address both of these questions.

CHAPTER 7

The LOCHWISP algorithm

Introduction

In the preceding chapter we outlined a mathematical basis which might allow us to solve two 'new' location problems on a plane, namely how to locate centres:

- (a) to maximize use;
- (b) to maximize the net balance of certain costs and benefits to users.

These problems are new in the sense that the existing literature does not contain, as far as the author is aware, any discussion of their properties or how to solve them.

For either method to be successful, the equations concerned (6.10 and 6.11; 6.12 and 6.13) would generally need to converge, through reiteration, on progressively better solutions. The existing literature on locational analysis, however, does not provide any mathematical analysis of their properties of convergence nor does any computational experience with these or similar equations appear to have been published. Hence it was not possible to know in advance whether these methods would succeed. The principal aim of the present chapter is therefore to examine how well the two methods worked in practice, i.e. we are mainly concerned with presenting computational experience.

We can ask two basic questions about each method:

- (a) will it be successful as a means of locating one centre?
- (b) will it be successful as the location stage of a location/allocation algorithm for locating several centres?

If Thiessen catchment areas are used in the latter type of algorithm, the distance between any demand point and the centre to which it is assigned can only be reduced during the allocation stage, as we noted in Chapter 6. Hence the number of trips emanating from each demand point can only increase and the associated travel costs must fall. Therefore the level of overall use and the associated net balance of benefits and costs, NBU, can only increase during the allocation stage. It follows that if either of the two equations produce better solutions during the location stage, then the whole algorithm should obtain progressively better solutions for that objective. In short, if the method works for one centre, an algorithm using it as the location stage to place several centres should also work.

This conclusion relies on inferences made from the distance minimizing properties of Thiessen polygons. If we allow catchment areas to overlap then we can no longer make such inferences and it becomes uncertain whether the allocation stage, however formulated, would lead to improvements in use and the associated travel cost. In this context the allocation and location stages might not work in concert, so without experience, we cannot say whether the whole algorithm would produce convergent results or not. The new algorithm written to test both methods did allow catchment areas to overlap, partly in the hope that it might be able to shed some light on this question.

Using iterative equations 6.10 - 6.13 a program called SEARCH was written to locate one centre so as to maximize use or net benefit. In trials with various starting positions a case of divergence was never encountered with either objective; these

tests were carried out with a small hypothetical data set and with population data for Edinburgh. Thus, from any starting position, SEARCH always moved the centre to locations with higher values of TRIPS or NBU, depending on which goal had been selected.

Of course, with high values of b local optima became relatively more frequent. The properties of the searching procedures, such as their efficiency and frequency of local optima can, however, be discussed more conveniently in a later part of the present chapter. The basic success of SEARCH meant that an algorithm to locate several centres according to these goals could be written with some confidence.

After the iterative equation for NBU had been derived and a few successful trials had been made with it, a mathematician was asked whether there might be any conditions in which divergence would occur. The reply was that convergence depended on a two by two matrix of derivatives at the point of solution itself. However, since it would be very difficult to establish these properties when the solution was unknown, the best policy would be to 'suck it and see'. Accordingly, it was thought best to continue using the method and see if any difficulties occurred in practice. If some cases of divergence had been encountered it might have been necessary to investigate the mathematical conditions for convergence more closely. To the author's relief, a case of divergence has not so far occurred.

Experience with NORLOC and LAP

This new algorithm was christened LOCHWISP (Location of Centres Heuristic with Iterative Search on a Plane). Though its main purpose has now been outlined, there were several other reasons why

it became desirable to write a new algorithm. Most of these stemmed from difficulties encountered with LAP and, since the general aim of thesis is to examine models and methods used in locating centres, it may be useful to give a brief account of experience with this well known package.

At the start of the research, one of the main aims was to make two of the main algorithms then in existence, NORLOC and LAP, operational at Edinburgh for empirical work on various services. There was some delay in making NORLOC operational, mainly due to the author's limited experience of computing at the time. After its first successful run, NORLOC presented very few further difficulties as a program, though the way in which it takes input data is so unusual that a special program was required to convert the census data to the peculiar format needed by NORLOC. However, once the limitations of the model on which NORLOC is based were realised, it seemed preferable to use the more complex and flexible LAP as the basis for locational work.

The most interesting feature of LAP was that it allowed centres to vary in attractiveness. If this option in the program is selected, during the location stage each centre is moved to the point within its catchment at which the expression

$$\sum_{i=1}^n p_i w_j / d_{ij}^b$$

(where w_j is the attraction of the centre in the catchment under consideration)

is maximized. The influence this goal has on the solution will obviously depend on the power assigned to distance, but it will tend to place the centre nearer dense concentrations of population

than a p -median solution would. It must be emphasized that the demand emanating from each point is still independent of distance, so the underlying model is not based on elastic demand. Rather, centres are located within their Thiessen catchments to maximize the above expression which is based on the gravity model.

Initially it was not entirely clear from the Iowa monograph (Rushton, Goodchild and Ostresh, 1973) whether the model was based on elastic demand or not and the point had to be confirmed in correspondence with the author of the program, Michael Goodchild.

Other attractions of LAP include its ability to solve the 'constrained' location problem where there are capacity constraints on the facilities. In addition LAP allows barriers to movement to be specified so that travel routes then have to go around the lakes, rivers and other topographic features concerned. All of these features made it desirable to have the algorithm available for use at Edinburgh.

The greater flexibility and complexity of LAP vis à vis NORLOC is probably one of the reasons why it proved more difficult to make the program operational here. In fact the program could not be compiled at first because the Edinburgh Fortran compiler on the ICL 4-75 was more fastidious about unassigned variables than the Fortran compiler at Waterloo where LAP had been developed. The modification made to LAP to remove this difficulty was later accepted by Michael Goodchild as a minor correction to be incorporated in future versions of the program.

The serious difficulty with LAP, however, was that when run on the ICL 4-75, it would search only when the starting positions of centres were left blank. In such cases it would then generate

random starting positions and proceed to search from there. If specific starting locations were supplied, it would evaluate the associated travel cost and stop without search. This was a serious drawback, because exploration of potential sites for swimming pools in Edinburgh required the use of specific starting positions, especially for solution of the additional facility problem. Furthermore, if we have no control over the starting positions, there is less scope for breaking out of local optima.

Despite the substantial amount of time spent in examining the program and in experimenting with the different arrangements in input, the reason why LAP behaved in this way on the 4-75 was never detected. Goodchild could not replicate the difficulty at Waterloo and eventually the quest to make LAP fully operational on the 4-75 was abandoned.

The latter difficulty with LAP gave further impetus to the construction of the new algorithm, since it was relatively easy to include the minimization of aggregate travel as one of the possible goals for the new program. The intention was therefore to use LOCHWISP to search from specific locations in the way we had intended to use LAP. Despite the difficulties encountered with LAP the experience of writing a new algorithm was surprisingly favourable on the whole. It is much easier for anyone to trace errors in a program he has written himself partly because each programmer has his own idiosyncracies. Moreover, any complex program in FORTRAN tends to be rather opaque to anyone other than its author.

LOCHWISP was developed on the ICL 2970 which came into service in the summer of 1978. After it had worked successfully on p-median problems, LAP was tried on the 2970, largely out of

idle curiosity. To the author's astonishment, it proceeded to search from the locations specified. It was not apparent why it behaved differently on the two machines. Experience on the 2970 thus only deepens the mystery, because the version which ran on the 2970 and the data file concerned were both copied directly from the versions used on the 4-75 without alteration of any kind.

Since it proved easy to move NORLOC to the 2970 as well, it was possible to compare the efficiency of the three programs in solving a standard problem in minimizing aggregate travel. The problem selected was that of locating seven centres, using the sites of Edinburgh's public swimming pools in 1971 as starting positions for search. The results of this comparison will be discussed later as part of a general comparison of the three programs.

The structure of the LOCHWISP algorithm

The basic structure of LOCHWISP is indicated in Figure 7.1. Although this flow diagram does not always exactly reflect the organisation of the Fortran program, it is probably as close as simplicity allows. From the main program various subroutines are called according to the nature of the problem and options selected. The two most important subroutines are CATCH, which defines the catchment areas (the allocation stage) and the more complex EVALOC which moves the centres within their catchment areas and evaluates the appropriate objective functions (the location stage). If overlapping catchment areas are used, the subroutine SHARE is invoked to allocate demand between the various centres, using a very simple spatial interaction model. Since the LOCHWISP program is fundamental to the rest of the thesis and may be of interest to

Figure 7.1 Structure of LOCHWISP program

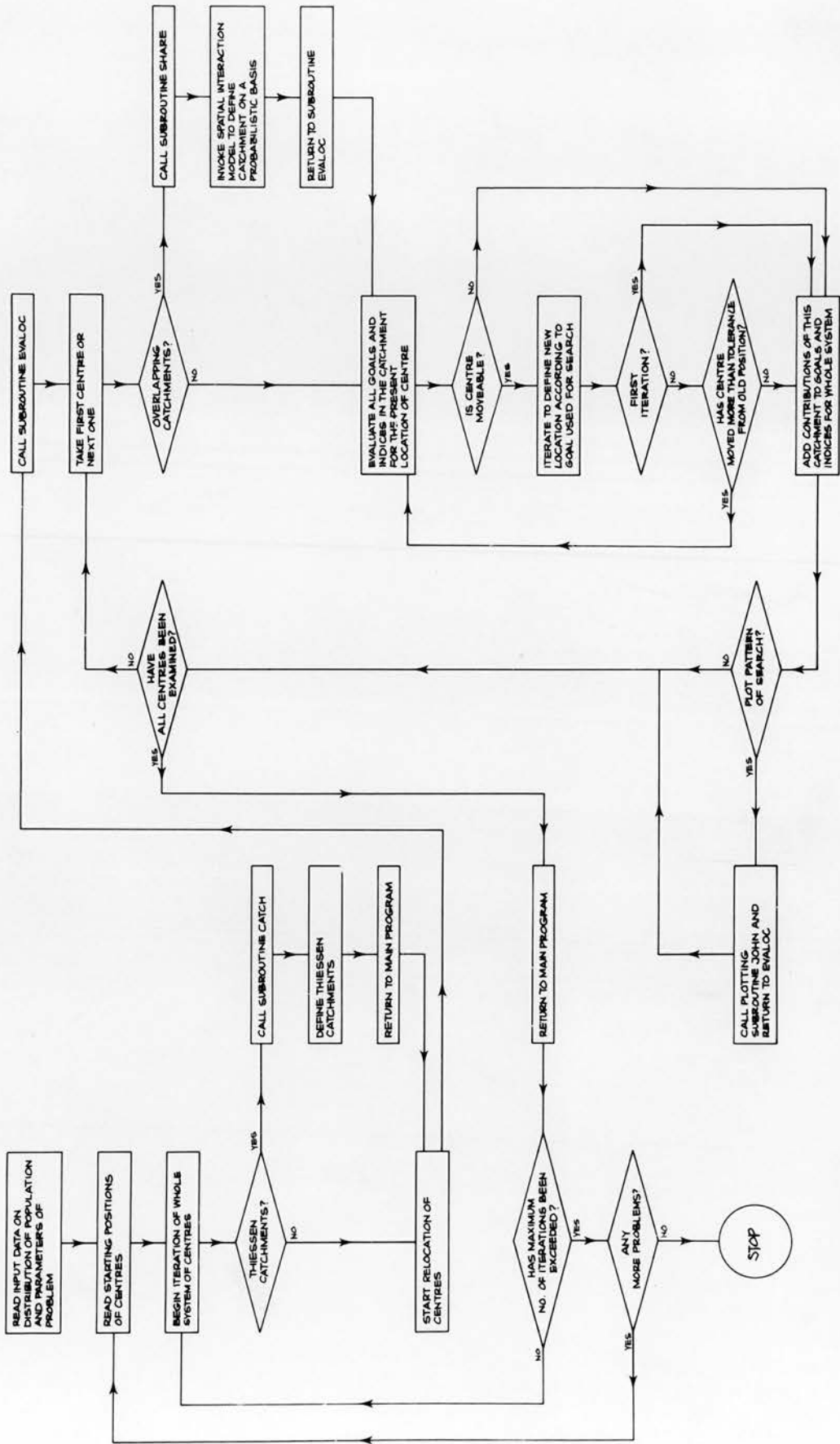


Figure 7.1 Structure of LOCHWISP program

other workers, the text of the main program and these three subroutines is given in Appendix 4. Another subroutine called JOHN was used to map the pattern of search but it is of less general interest and is therefore not included in the appendix.

Since an algorithm which can treat overlapping catchments in its search has not been found in the literature, the way in which LOCHWISP treats such catchments may be of particular interest. If catchments overlap, the subroutine defining Thiessen catchments is bypassed; during the relocation subroutine, EVALOC, each centre's catchment is then notionally defined as the whole city. If successive centres were then located within that catchment on the same basis as within the Thiessen catchments, for relatively inelastic problems each of the m facilities would tend to be moved to the city centre. LOCHWISP would, in effect, merely be solving the same location problem m times, partly because there is then no inbuilt tendency for the centres to repel each other.

To circumvent this difficulty, in subroutine SHARE the potential demand from each grid cell is split between various centres using a spatial interaction model of the form

$$T_{ij} = p_i e^{-b_s d_{ij}} / \sum_j e^{-b_s d_{ij}} \quad (7.1)$$

where b_s defines the rate at which the attraction of j declines with distance. The proportion of potential demand at i which is assigned to centre j then depends on the attraction of j relative to that of all the centres where the attraction of each centre diminishes exponentially with its distance from a demand point according to b_s . Thus, if there are several centres closer to i than j , very little of i 's demand will go to j . Each centre then has a catchment resembling an irregularly shaped 'cone' whose sides may fall sharply near rival centres. If all centres were located near the city centre, they would share demand from each peripheral cell fairly equally. A later

modification to the model allows the centres themselves to vary in attraction. Nevertheless, in order to examine the purely spatial aspects of LOCHWISP's searching behaviour it is simpler to postpone discussion of this factor.

With its own catchment defined, a centre is then treated as it would be in a Thiessen catchment and relocated according to the goal selected. To maximize its own demand, a centre will then tend to be repelled from the locations of other centres, but less strongly than with Thiessen catchments.

LOCHWISP evolved in stages from an earlier program called GEANS which simply evaluated various measures of accessibility for a given set of locations without searching to improve it. Partly for this reason, it has a certain amount of redundant program code and hence could be streamlined. Another reason for this redundancy is that transparency in the program code has often been preferred to conciseness.

Each centre is designated as fixed (1) or moveable (0). One iteration of LOCHWISP is said to be complete when all the moveable centres have shifted at least once. During one of these iterations a particular centre will be moved repeatedly within its catchment area until the distance displaced is less than a specified tolerance value, TOL. Thus within one iteration of the whole algorithm a particular centre may move several times (Figure 7.1). The standard tolerance values used were .20 and .10 units. With a grid based on units of 500 metre squares, .20 is equivalent to 100 metres on the ground, an adequate degree of accuracy for selecting a good general location rather than a specific site.

In a typical run each centre might well move more than four or five times in an early iteration but only once in later iterations, as the pattern converged. The very first iteration is an exception

to this general rule since each centre is allowed to move only once during the course of it. This is merely a convenient device for evaluating the various objectives and indices of accessibility for the initial positions so that they can be compared with those for the final locations.

Although LOCHWISP can only search using one objective at a time, it does note the values of the other goals and indices as the search proceeds. For instance, when aggregate travel is being minimized, the level of overall demand for the specified value of b and the net balance of costs and benefits are recorded as well. A number of other objectives or indices which LOCHWISP is not able to use as goals for search are also noted: the tributary population of each catchment, the population within a specified covering radius and the maximum distance travelled by any user. All objectives and indices are computed for each catchment and for the whole system.

As noted earlier, the latter two indices can be taken as measures of equity. Since it does not seem to be possible to use a gradient type of search to optimize either goal, LOCHWISP cannot identify a set of locations which maximizes equity. However, the output from LOCHWISP does allow us to observe whether the other goals and indices improve when we search with a particular objective. Thus, given two solutions with similar levels of efficiency but different levels of equity, we could take that with the higher level of equity as the better solution. In computing NBU, it should be noted that a check is first made to see whether the value of a trip is less than its cost. If it is less, then in accordance with the argument outlined in Chapter 3, the trip is not made as far as the computation of NBU is concerned.

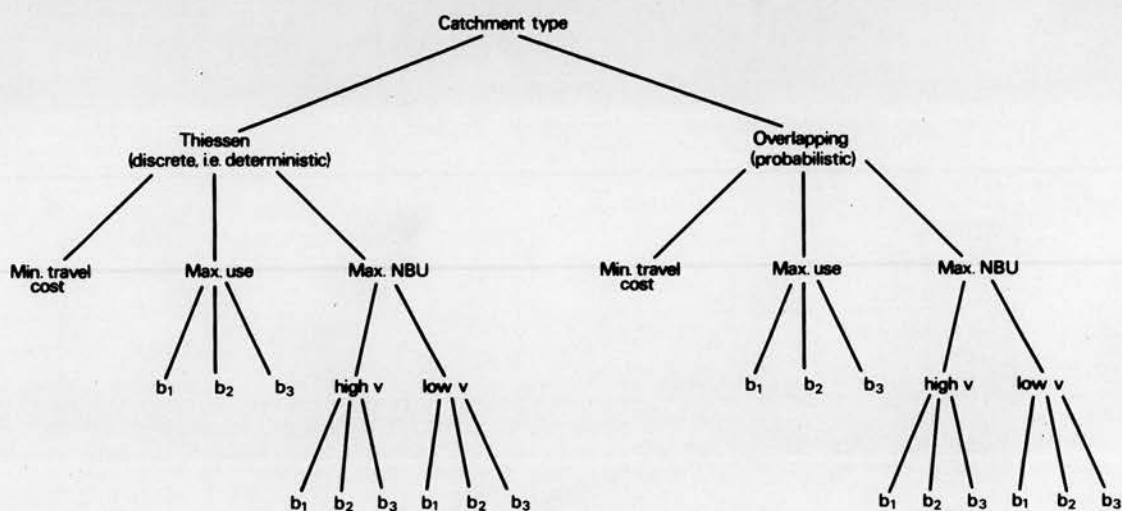
Range of problems covered by LOCHWISP

The family of locational problems which LOCHWISP is able to solve can be described as a tree (Figure 7.2). The tree first divides according to whether we assume that users will only come from within the Thiessen polygon around a centre or whether they may potentially come from anywhere in the city. Thereafter each branch splits according to which of the three goals is pursued. When demand is elastic we can specify a range of b values from the relatively inelastic value of $b = 0.125$ to the more elastic values of $b = .25$ and $b = .50$. This set of b values has been shown respectively as b_1 , b_2 and b_3 in Figure 7.2 to indicate the range of elasticity subsumed under the objectives of maximizing demand and net benefit. A potential range of values for v is also alluded to through notationally high and low values.

As Figure 7.2 illustrates, if we take three values of b and two values of v , using only Thiessen catchments, we have ten forms of the location problem to explore for a given number of centres. Moreover, if we suspect there may be local optima and use, say, five different sets of starting locations for each of the ten problems, we then potentially have fifty different solutions to examine. With a further fifty from overlapping catchments we could then have a hundred solutions for one value of m .

Fortunately this daunting multiplicity is less difficult in practice. When demand is relatively inelastic, local optima are less numerous. Furthermore, it makes little sense to assign a high value, v , to use of the service and at the same time make demand very elastic; conversely if v has a low value, then it seems irrational to make b relatively inelastic. This further

FIG 7.2 THE FAMILY OF LOCATION PROBLEMS WHICH LOCHWISP CAN SOLVE



reduces the number of solutions which are of real interest.

Logically, the values for v and b should be linked, but a satisfactory way to do this has not been found within the present framework and we therefore have to examine v and b in discrete combinations. If we knew from empirical work on the particular service under consideration how elastic demand actually was, most of the range of values could be thereby eliminated. As noted in Chapter 3, however, there are very few studies sufficiently rigorous to yield estimates of b . In the case of vital medical and fire services, we can assume that demand is inelastic on common sense grounds. For libraries and swimming pools we can guess that demand is moderately elastic but there is no empirical work to substantiate the guess and we therefore have to explore a range of b values.

Even with such reductions as can reasonably be made with swimming pools and libraries in mind, it is impossible to explore every one of the large number of possibilities which remain. All we can do is take a fairly representative sample from the family of solutions. In practice, this was less of a problem than had been feared initially because different goals and values of b sometimes produced very similar sets of final locations.

Testing LOCHWISP

It has been possible to check LOCHWISP to a limited extent by comparing the results with those derived from NORLOC and LAP, but only for the p -median problem, since LAP and NORLOC do not solve the other problems. Aggregate travel to the seven public swimming pools in Edinburgh in 1971, as computed by LOCHWISP is 2,121,837 units of cost, which is identical to the answer from LAP, though the latter

is given only to four significant digits. When these locations are used as the starting points for search, LAP and LOCHWISP produce very similar final solutions (Table 7.1).

The very slight differences between the solutions from LAP and LOCHWISP are probably a result of two factors. First LOCHWISP only allows each centre to move once during its very first iteration, whereas in LAP's first iteration they continue to move till less than the distance tolerance before their catchments are reallocated. The trajectory of search thereafter will inevitably be slightly different. Second, the two programs have slightly different ways of applying the distance tolerance. A distance tolerance of .10 in LAP means that search is continued if the displacement either of the x-co-ordinate or the y-co-ordinate is greater than .10. In LOCHWISP the search continues until the actual distance the centre moves is less than the tolerance. This difference may produce fine variations in the paths of search traced out by the two programs from the same starting position. Nevertheless, the differences between the LAP and LOCHWISP solutions are so small that the aggregate travel associated with each differs by less than one per cent and in another trial was identical to the fourth significant digit. We can therefore safely say that this comparison with LAP confirms the equivalent results from LOCHWISP.

The solution obtained from NORLOC, at first appears slightly different from the other two. However, closer inspection shows that five of the seven final locations are virtually identical to the corresponding locations for LOCHWISP. In addition NORLOC finally locates the seventh centre at a position quite close to that reached by the third centre sited by LOCHWISP and LAP. The

TABLE 7.1

Comparison of solutions by LAP, LOCHWISP and NORLOC
to a standard problem with seven centres

<u>Centre</u>	<u>Initial</u>		<u>Final Positions</u>					
	<u>Position</u>		<u>LAP</u>		<u>LOCHWISP</u>		<u>NORLOC</u>	
1.	16.2	12.6	8.63	10.82	8.58	10.75	8.70	10.85
2.	17.2	16.7	12.36	18.62	12.37	18.62	12.45	18.45
3.	20.2	13.4	17.84	12.24	18.41	12.46	14.95	5.65
4.	21.4	19.0	21.26	17.65	21.39	17.60	21.40	17.50
5.	29.9	14.8	27.81	12.89	27.94	12.65	27.90	12.55
6.	21.6	11.8	24.49	5.78	24.47	5.77	24.35	6.05
7.	18.6	10.9	15.13	6.58	15.20	7.68	17.35	11.90
<u>Travel cost</u>			1,195,000		1,205,767.0		1,192,500	
<u>Time of Computation</u> (CPU secs)			12.27		7.40		207.27	
<u>No. of Iterations</u>			10		11		10	
<u>Distance Tolerance</u>			.10		.10		.25	

biggest discrepancy is for the final position of the third centre with NORLOC which is over two grid squares from the nearest centre of the LOCHWISP solution. Nevertheless, the broad similarity of NORLOC's solution tends to support the argument in Chapter 6 that NORLOC's discrete search would tend to find the same minimum for any given problem as a search based on partial derivatives.

It is interesting that the travel cost for NORLOC's solution is slightly lower than the other two. It is conceivable that NORLOC's thorough plodding search may sometimes find slightly better solutions than the other two programs. If LOCHWISP is allowed to continue searching after the eleventh iteration, however, it finds a solution of 1,192,278 units of travel cost on the twentieth iteration, using a total of 9.22 secs of CPU time. Thus the main reason for NORLOC's lower solution in Table 7.1 must be that it has simply spent more time searching at that stage.

It is also interesting to compare the times of computation of the three programs (Table 7.1). In each case the source program had been compiled with optimisation parameters in order to reduce time of execution. The striking feature of the table is that LAP and LOCHWISP, based on iterative search, are both over fifteen times faster than NORLOC with its step by step exploration. In fact the difference in efficiency is even greater because the precision of NORLOC's solution was only to within 0.25 units of the grid (or 125 metres) whereas for LAP and LOCHWISP it was roughly 0.1 (or 50 metres).

Surprisingly, LOCHWISP took even less time than LAP, needing 7.40 secs of CPU time as opposed to 12.27 secs. The reason may

be that LAP has a more complex structure reflecting its ability to solve a different range of problems from LOCHWISP. For instance it includes a linear programming routine for use with the 'constrained' location problem. On the other hand LOCHWISP computes a wider range of indices of accessibility; it was written to be easy to check rather than to run efficiently, so it has some features which could be further streamlined; and it has a few options not included in LAP. Moreover, on the trial run the options selected were basically the same on each. It may be worth adding that its time of computation on this trial problem was consistent with the time needed to solve a variety of other problems with the same number of centres and demand points, so that this particular problem did not appear to be an unusually favourable one for LOCHWISP. Furthermore the parameters used to control the runs, the maximum number of iterations and the distance tolerance, were chosen to be slightly unfavourable to LOCHWISP. A more complete comparison of the two programs on a wider variety of problems could be undertaken, but our main aim here is merely to establish that LOCHWISP is a reasonably efficient program. Clearly, in any research where the real cost of computation has to be considered, an iterative type of algorithm will be preferred to NORLOC.

As the Lund monograph (Törnqvist, Nordbeck, Rystedt and Gould, 1971) indicates, in NORLOC all the population in a grid cell, is assumed to be located at the centre. Thus, when we specify that a number of people are located in cell (5,8), NORLOC in effect treats their location as (5.5,8.5). However, a facility with co-ordinates (5,8) is treated as if it were at those exact

co-ordinates. When an option was introduced in LOCHWISP to allow the locations of population and facilities to be treated in a manner equivalent to NORLOC, aggregate travel to seven stationary centres as computed by the two programs was found to be identical.

Basically LOCHWISP, like LAP, treats the location of population and facilities on the same basis. Since the grid of population data had been originally set up for use with NORLOC it was therefore necessary to add 0.5 to all population co-ordinates read into LOCHWISP and LAP to ensure consistency and comparability. Thus the population cell read into NORLOC as (8,5) had to be read into LOCHWISP as (8.5,5.5). The results presented in Table 7.1 were computed on that basis.

Further checking of LOCHWISP against NORLOC brought to light some features of the latter which might have otherwise gone undetected and do not appear in the LUND monograph. Checking was first done using locations for pools which were rounded to a whole number by assigning each pool to the S.W. corner of the cell in which it was located. With these locations the hinterland populations or capacities of several centres computed by LOCHWISP were substantially different from those computed by NORLOC when no search was requested. When LOCHWISP was modified to check whether demand points were equidistant from two or more facilities, 28 out of 342 demand points were found to be equidistant. Inspection of NORLOC's program code revealed that when a demand point is equidistant, the relevant Fortran IF statement implicitly assigns it to the second centre, i.e. the centre with the higher numerical label. On the other hand LOCHWISP assigns the demand to the first of the two centres. When the appropriate statement

in LOCHWISP was altered to match NORLOC, the capacities computed by the two programs were then identical.

Since decision makers might regard the population in the hinterland of a site under consideration as quite significant, it is desirable that the literature on a location/allocation algorithm should state explicitly how equidistant centres are treated. It should also give the user the option of assigning demand to the first of the equidistant centres and then to the second so that he can assess what difference it makes.

Better still, the program could split demand equally between equidistant centres. In fact a subroutine called ADJUST was written to do that for LOCHWISP, but it proved unsatisfactory with three or more equidistant centres, a situation not unknown with a NORLOC type of grid and facility locations based on whole numbers. A more complex subroutine could have been written to handle such cases but it was not felt to be worthwhile in practical terms because a simpler solution was found.

When the locations of the facilities were given more precisely, this problem disappeared and capacities computed by NORLOC and LOCHWISP were identical. Since the subroutine ADJUST was very rarely used thereafter it is not shown in Figure 7.1 and has not been included in Appendix 4. When the locations of facilities are given to the second decimal place, it is very unlikely that a demand point will be equidistant from two centres; it is then practically impossible for three equalities to occur. Apart from the need for algorithms to be more explicit on this point, the main conclusion drawn from this experience was that the populations of the hinterlands derived from NORLOC should be

treated with caution, when facility locations are given in whole numbers.

NORLOC was originally used to check LOCHWISP because LAP did not search from specified locations when run on the ICL 4-75. To make LOCHWISP equivalent to NORLOC, originally 0.5 was added to population co-ordinates when distance was being computed from a demand point to its nearest centre. The first results obtained for search with LOCHWISP in this framework were puzzling. Initially the iterative search would reduce travel costs to a relatively good solution; thereafter further iterations would gradually increase travel cost by small amounts. This experience was repeated with the goals of maximizing use and net benefit. However, when distance was computed without adding 0.5 this problem disappeared; successive iterations then improved the objective function by ever decreasing amounts in all three cases, as desired.

The reason for this difficulty is probably that the constant added to the distance expression alters the partial derivative. The differentials used in search are then no longer appropriate to the travel cost function. Thus the addition of a constant (here 0.5) to the distance expression does not cause any difficulties with NORLOC's type of search but it does with the gradient type of search used by LOCHWISP.

Ironically the use of NORLOC to check LOCHWISP brought to light more problems in NORLOC than in LOCHWISP; but it also drew attention to the need to document some features of location/allocation algorithms more explicitly.

Searching properties of LOCHWISP

(a) The location of one centre

The main purpose of writing LOCHWISP was to demonstrate that the goals of maximizing use and net benefit could form the basis for an alternating type of algorithm. It has already been argued that if the iterative procedures work in locating one centre then it necessarily follows that an alternating algorithm to locate several centres using these procedures must also work as long as catchment areas are of the Thiessen type. Another aim was to investigate the searching properties of such an algorithm for both Thiessen and overlapping catchment areas.

Whereas kilometres were notionally used to measure distance in Chapters 3 and 4, the grid framework used to describe the distribution of population in Edinburgh was based on units of 500 metres. Thus distances in this grid have twice the numerical size of the preceding ones. Keeping the same numerical value of b , a distance decay effect of e^{-bd} now becomes e^{-2bd} ; the grid therefore has the same effect as doubling the value of b . In real terms the demand cone formerly obtained from $b = .25$ now has the same profile as that for $b = .125$. Therefore instead of the earlier series of b values (.25, .50, .75 and 1.0), we now use .125, .25 and .50, dropping the third value for the sake of simplicity.

The easiest way to study the searching properties of LOCHWISP is to try to find the best location for one centre under various objectives from a variety of starting positions. Accordingly, four starting positions were selected; to the S.West; S.East; N.West; and N.East of Edinburgh. If a starting position has

exactly the same location as a demand point, a division by zero occurs in computing the derivatives of travel cost (6.8 and 6.9). To avoid this difficulty starting positions are offset a small distance from points of demand. For the goal of maximizing trips, only the extreme values of b , .125 and .50, were used.

Similarly in maximizing NBU, analysis was confined to a problem with a high value of v and low elasticity ($v = 10.0$ and $b = .125$) and a problem with a lower value of v and high elasticity, ($v = 5.0$ and $b = .50$). Because of the change in distance units these are equivalent to values of 5 and 2.5, if v is measured in kilometres.

The starting positions and solutions to these test problems are shown in Figure 7.3. For the three problems where elasticity is either low or absent the search invariably finds its way to essentially the same location in the south west of the city centre at Tollcross, no matter where it starts. This suggests that the surfaces being searched for each of the three goals are probably fairly simple ones with only one optimum. Thus by using the gradient to guide its search, LOCHWISP is able to reach that single optimum fairly quickly and directly. However, when elasticity is high ($b = .50$; $v = 5.0$ with $b = .50$) search terminates in solutions which can be shown to be only locally optimal (Table 7.2 (c) and (d)). With high values of b there are likely to be many local peaks on the surfaces of both TRIPS and NBU. Thus by ascending the contours around the starting position LOCHWISP only reaches a local summit and of course stops there.

A striking property of the search is the length of the first step. In many cases more than half the distance to the optimum is travelled in the first iteration, irrespective of whether the

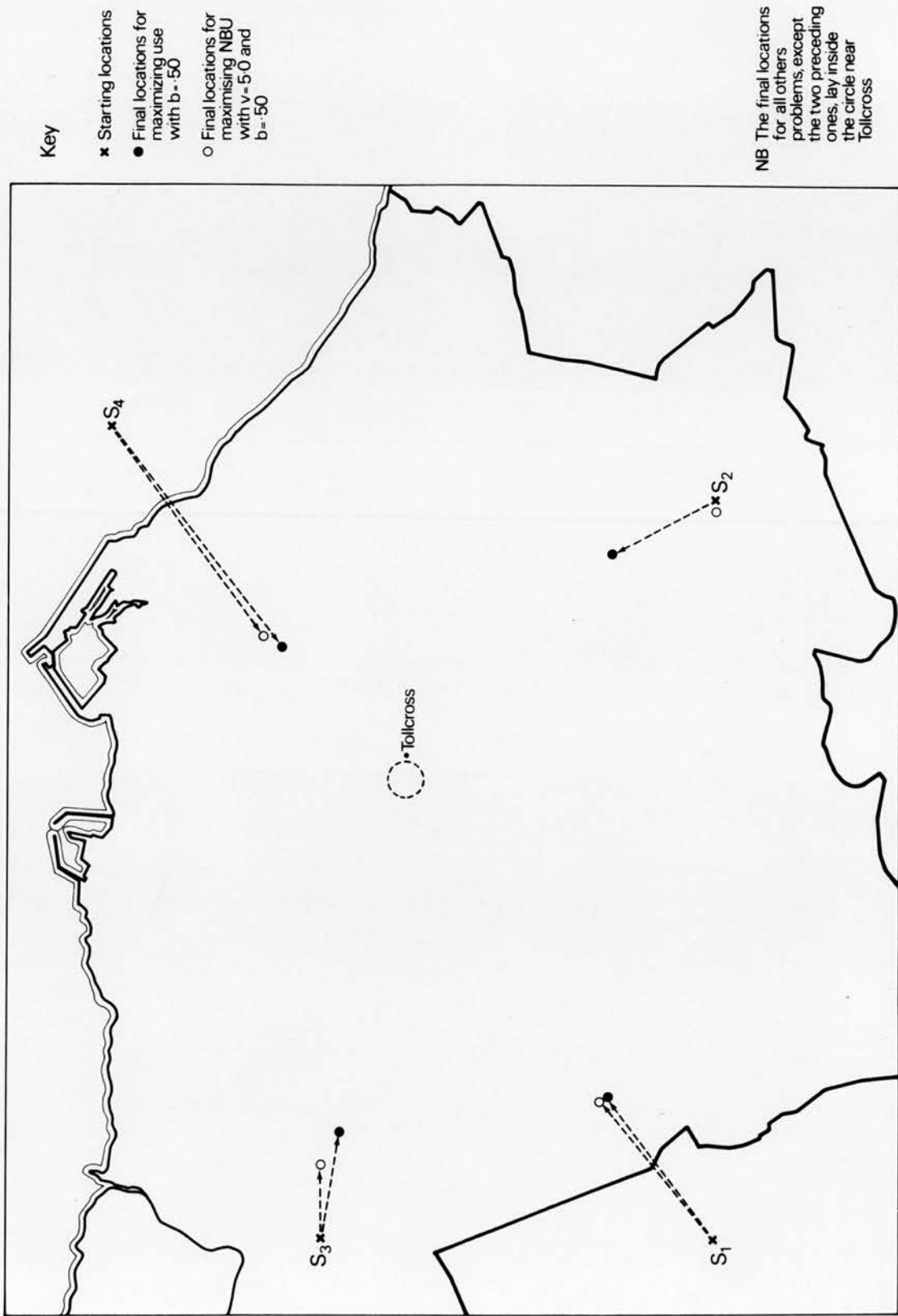


Figure 7.3 Results of locating one centre in Edinburgh according to various objectives, using different starting locations

TABLE 7.2

Characteristics of search for location of one centre
in Edinburgh by LOCHWISP

(a) Minimizing travel cost

<u>Starting location</u>	<u>Final location</u>	<u>Distance moved in iteration 1 as percentage of total distance moved</u>	<u>Number of iterations needed (to reach toler- ance)</u>	<u>Solution global or local</u>
1. 4.1 4.1	17.0 12.6	82.3	7	global
2. 25.1 4.1	17.3 12.6	49.5	7	global
3. 4.1 15.1	17.0 12.7	69.5	7	global
4. 27.1 21.1	17.3 12.8	85.3	6	global
5. 20.0 35.0	17.2 12.9	97.4	4	global

(b) Maximizing use (b = .125)

<u>Starting location</u>	<u>Final location</u>	<u>Distance moved in iteration 1 as percentage of total distance moved</u>	<u>Number of iterations needed (to reach toler- ance)</u>	<u>Solution global or local</u>
1. 4.1 4.1	16.9 12.7	56.6	13	global
2. 25.1 4.1	17.5 12.7	19.2	12	global
3. 4.1 15.1	16.9 12.7	38.9	12	global
4. 27.1 21.1	17.4 13.0	66.3	12	global

(c) Maximizing use (b = .50)

<u>Starting location</u>	<u>Final location</u>	<u>Distance moved in iteration 1 as percentage of total distance moved</u>	<u>Number of iterations needed (to reach toler- ance)</u>	<u>Solution global or local</u>
1. 4.1 4.1	8.0 8.0	77.7	4	local
2. 25.1 4.1	23.7 7.0	8.0	18	local
3. 4.1 15.1	7.1 14.7	54.5	7	local
4. 27.1 21.1	20.9 16.4	75.0	9	may be global

(d) Maximizing NBU (v = 5.0; b = .50)

<u>Starting location</u>	<u>Final location</u>	<u>Distance moved in iteration 1 as percentage of total distance moved</u>	<u>Number of iterations needed (to reach toler- ance)</u>	<u>Solution global or local</u>
1. 4.1 4.1	8.0 8.0	40.0	9	local
2. 25.1 4.1	25.0 4.1	99.9	2	local
3. 4.1 15.1	6.1 15.1	55.4	4	local
4. 27.1 21.1	21.0 16.9	8.8	17	may be global

starting position is near the optimum or not. In particular, for minimizing travel cost the first step is generally longer than for the other objectives (Table 7.2), exceeding 80% of the distance from initial to final position in three cases. A plot of the intermediate positions shows that the route to the optimum for this objective is often close to a straight line; for the other objectives the route is usually reasonably straight. Thus for all three objectives a searching strategy based on partial derivatives is very efficient because it moves fairly directly towards the optimum and frequently gets quite close to it after only two or three iterations. This is the main reason why LAP and LOCHWISP are more efficient than NORLOC. An even more striking illustration of the efficiency of the method is the fact that from the very distant starting position of (20.0, 35.00), located almost on the Fife shore of the Firth of Forth, 97% of the distance to the point of minimum travel cost is traversed on the first iteration, as Table 7.2 (a) shows. Experiments with even more distant starting locations obtained even longer first steps. As a general rule, the more eccentric the starting position in relation to the catchment then the longer proportionally is the first step.

except one

In all cases/the solution based on net benefit was very close to the solution which maximized use with the same value of b . This is probably because the most sensitive element in NBU is the first term which is simply the number of trips multiplied by the constant v (see equation 3.16). Thus the first term dominates the value of NBU and so produces a solution similar to that for maximizing trips. Further comparison shows that the solution for NBU is generally nearer the starting position than the

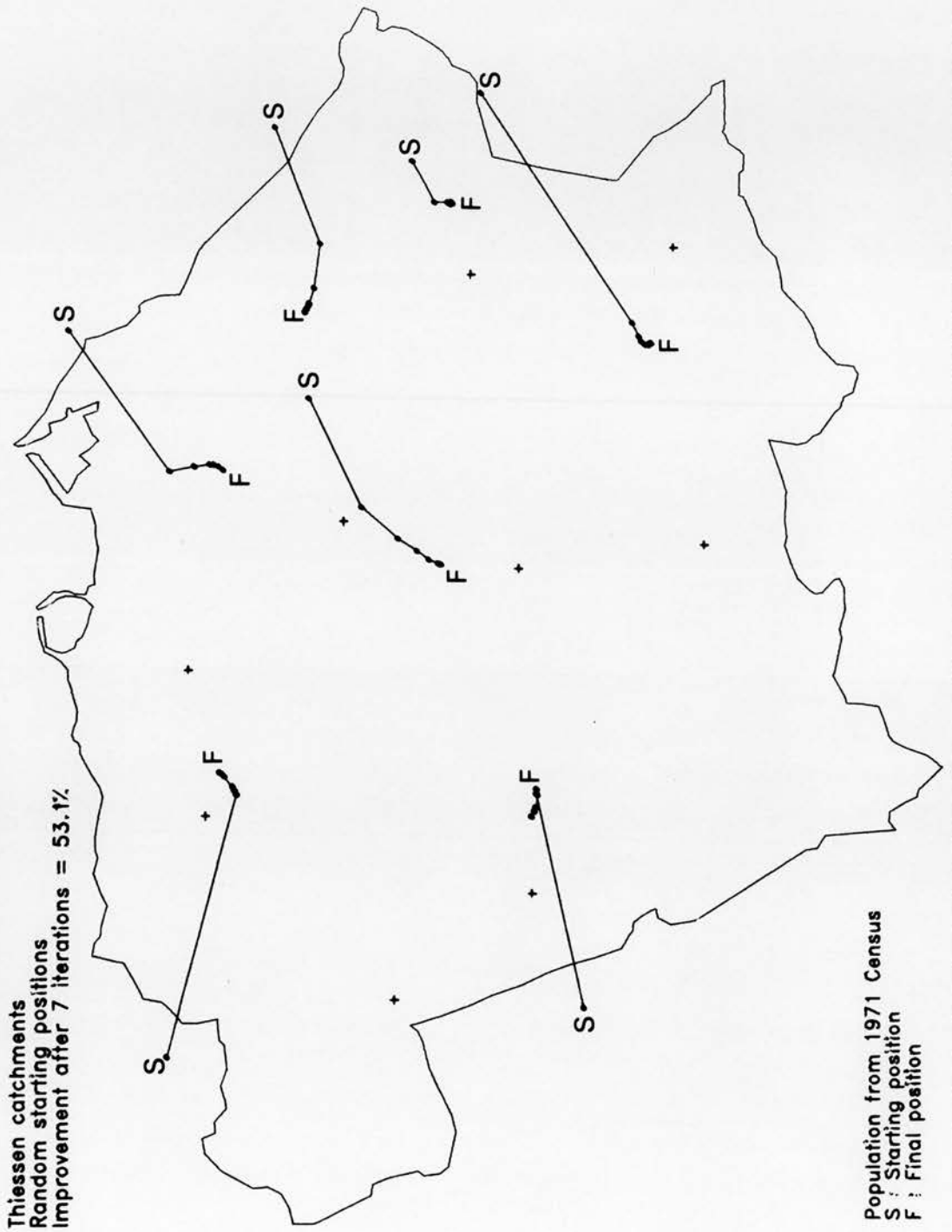
equivalent use maximizing solution. It is not completely clear why this happens but it may be that NBU stops more quickly in a local concentration of population because it has two terms both of which are improved by locating near an important concentration of people. In relation to a nearby cell with a particularly large population, NBU therefore counts distance twice in a sense and may be more prone to stick there. In consequence the use of NBU discounts outlying or dispersed population to an even greater extent than does the use maximizing solution and so provides an even less equitable solution.

(b) The location of several centres

To enable the searching movement of several centres to be plotted for examination a subroutine called JOHN was added to LOCHWISP. Basically, this subroutine allowed the lines of search to be mapped using the GIMMS computer mapping system, developed by Tom Waugh of the Department of Geography, University of Edinburgh. The idea for this way of plotting the results and the core of the subroutine itself were provided by John McCalden, a postgraduate student in Geography at Edinburgh.

To examine LOCHWISP's pattern of search for several centres a set of random starting positions was used with a problem concerning seven facilities. When the seven centres are located to minimize travel cost it is again striking how relatively straight the path to the optimum is in most cases and how relatively long the first step in the path is (Figure 7.4), especially when the starting position is eccentric to the potential catchment. When we recall that each centre moves only once during the first iteration of LOCHWISP whereas the distance between the second and

Figure 7.4 : Lines of search for location of 7 centres in Edinburgh
to minimize travel cost

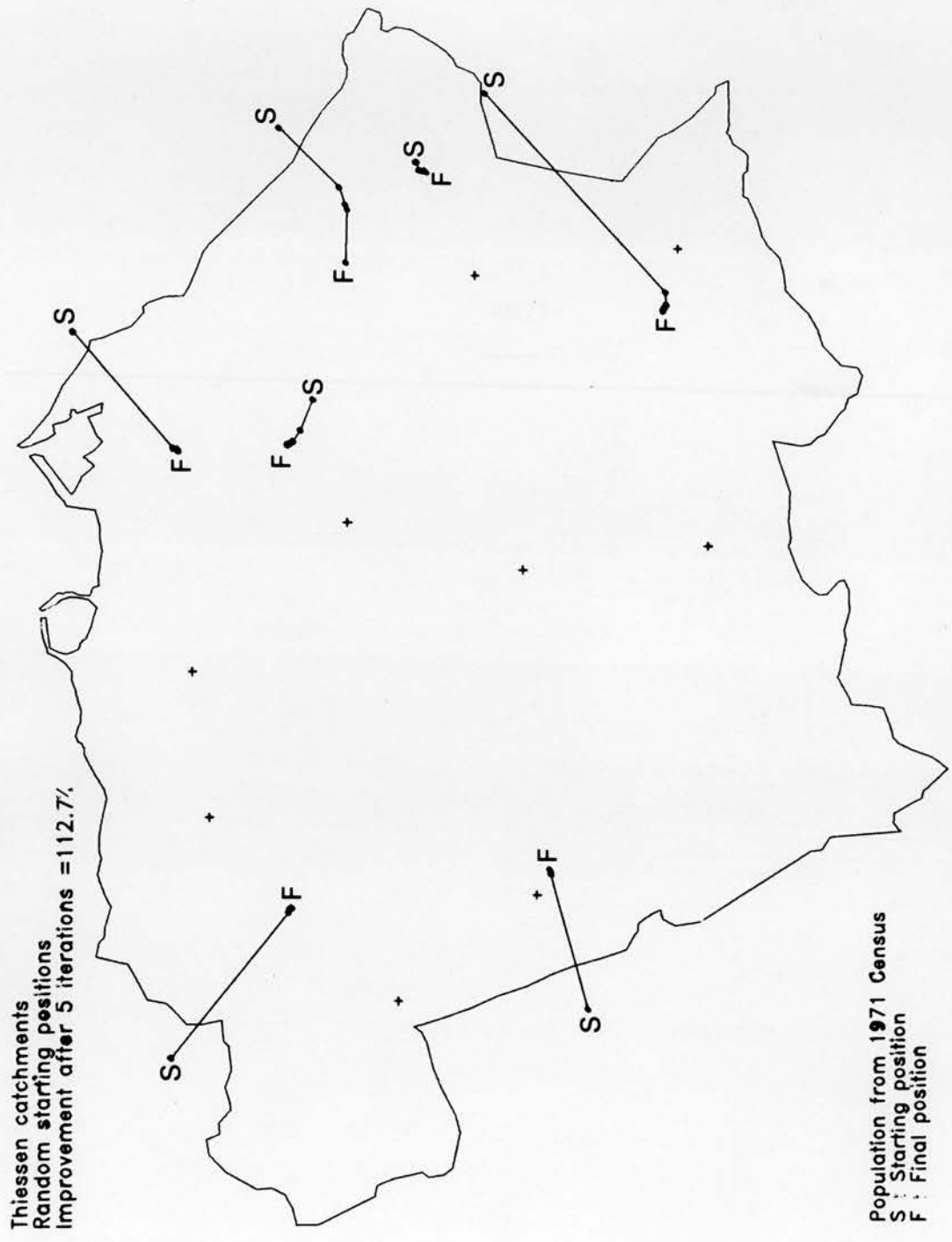


third positions of a centre may span several moves, the size of the first step is even more impressive.

When the seven centres are located to maximise use with the same starting positions, the lines of search and the final positions for $b = .125$ and $b = .250$ are both virtually the same as for minimizing travel cost and have not therefore been shown. When $b = .50$, six of the centres move in the same direction as before but do not travel quite so far. However, one facility, that starting near the centre of Edinburgh and a little to the east of it, travels a short distance north west (Figure 7.5) whereas it moved a much longer distance south west to the vicinity of Bruntsfield when b was $.125$ and $.250$. Though the large concentration of population in the Bruntsfield-Gorgie area lies within the catchment of this centre's initial position, it is on the edge of the catchment. With the lower values of b the concentration in Bruntsfield-Gorgie is large enough to attract the centre but when b is $.50$, its attraction is attenuated so much by the distance that a smaller but nearer concentration of population in the vicinity of Leith Walk exerts a more powerful pull. Most of the facilities, however, move almost as far when demand is inelastic. Thus, within the smaller catchments created when $m = 7$, even with a b value of $.50$, the tendency to stick in a local optimum is not as great as might have been expected from experience with $m = 1$.

A plot of the search based on NBU with $v = 10$ and $b = .125$ was very similar to that for minimizing travel cost and is therefore not shown. With $v = 10$ and $b = .25$ there was also little change, though a few facilities did not travel quite as far

Figure 7.5 : Lines of search for location of 7 centres in Edinburgh
to maximize use with $b=0.500$



along their path of movement as before. With $v = 10$ and $b = .50$ the plot was similar to that for maximizing use with the same value of b in that the location of the facility starting just east of the city centre was again moved north westwards instead of going to Bruntsfield (Figure 7.6). However, the location of the facility starting virtually on the eastern boundary of the city took a radically different path, moving northwards instead of south westwards. Thus the solution with $v = 10$ and $b = .50$ left the whole south eastern part of the city unserved, but assigned three centres out of the seven to the extreme eastern corner of the city defined broadly by the areas of Craigmillar, Craigentenny and Portobello. This result strongly reinforces the point that the goal of maximizing NBU yields even less equitable solutions than maximizing use and it also tends to end its search more often in purely local optima.

To investigate the effect of changing the value of v , solutions were also computed for the same range of b values and $v = 5.00$. When b was $.125$ the results were virtually identical. However, with $b = .25$ and $v = 5.0$ the final solution was more like that already obtained for $b = .50$ than that previously obtained for $b = .25$ (Figure 7.7). The results with $b = .50$ were very similar to the preceding ones, though the locations of the two facilities were moved slightly less far along their line of search. When solutions with $v = 2.5$ were plotted, they were found to be very similar to the lines of search for $v = 5.0$ but the distances travelled tended to be even shorter. In general, lowering the value of v tends to slightly shorten the distance travelled and to produce a slightly less equitable solution spatially. It thus has a similar effect to

Figure 7.6 : Lines of search for location of 7 centres in Edinburgh
to maximize NBU with $v=10.0$ and $b=0.500$

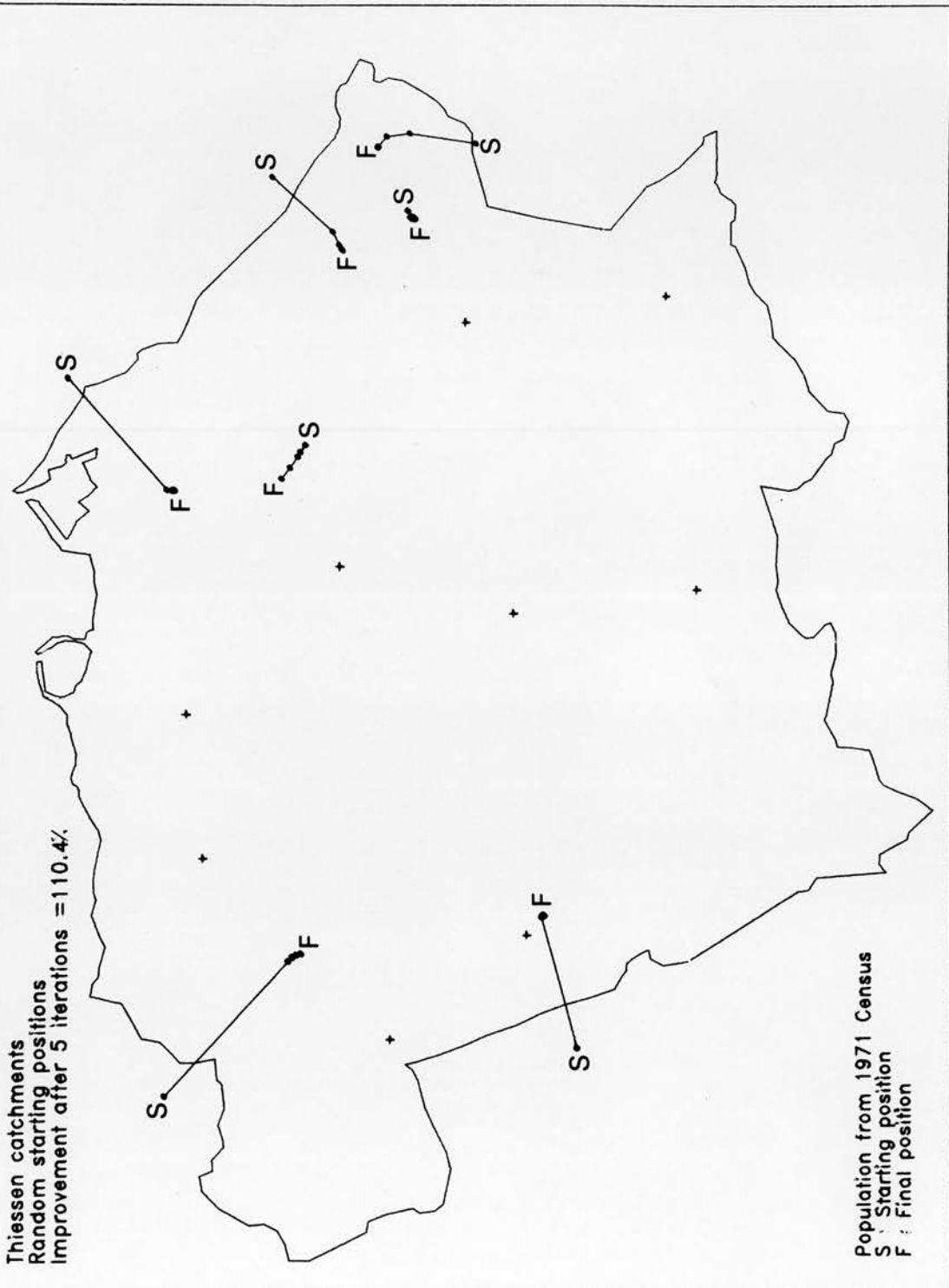
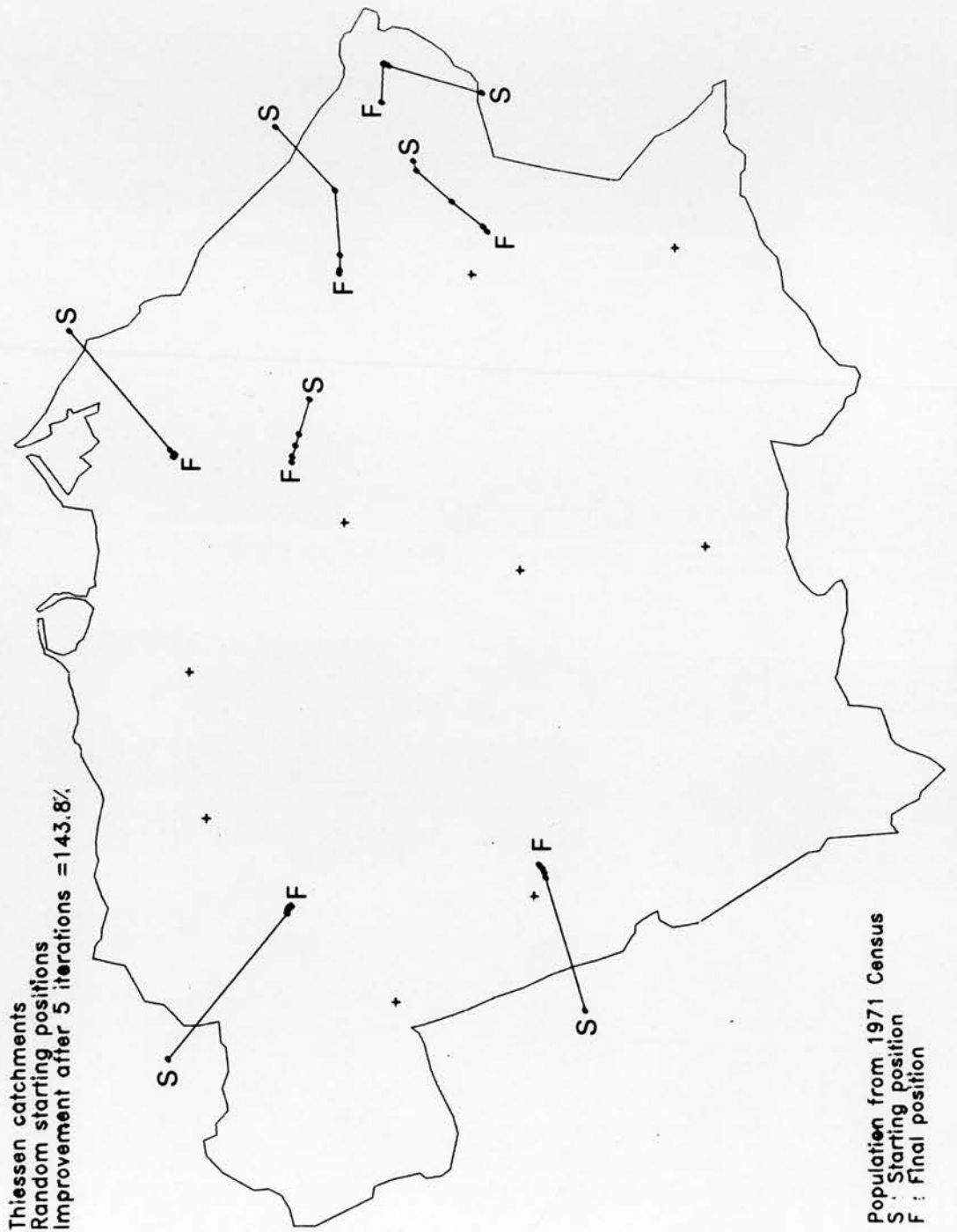


Figure 7.7 : Lines of search for location of 7 centres in Edinburgh
to maximize NBU with $v=5.0$ and $b=0.250$



increasing the value of b .

NBU is a more complex objective than travel cost and total level of use, since it depends on three parameters, v , b and c_t , the latter being the cost of travel per unit of distance. To reduce this complexity, we have used a value of 1 for c_t throughout the preceding analysis. If b and c_t are held constant, the larger the value of v then the more weight we give to the first of the two elements comprising NBU and therefore the closer we approximate to the goal of maximizing use. On the other hand if we make demand virtually inelastic by using a value of b close to zero the level of demand is almost the same at all locations. The second of NBU's terms (3.16) then dominates the choice of location, so we are then essentially minimizing travel cost. Thus the goals of minimizing travel cost and maximizing use can be seen as special cases of maximizing NBU. A corollary of this argument is that if b is close to zero, it hardly matters what value v has. In this sense b takes precedence over v in terms of its influence on NBU.

We can say little in this study about what the value of v should be. We can only explore the consequences of using relatively high and low values. In Chapters 3 and 4 we used values of 5 and 10 for v , equivalent to 10 and 20 in the grid framework. In retrospect this seemed rather high, especially in relation to swimming pools for which demand may be at least slightly elastic. This is the reason why the lower values of 10, 5 and occasionally 2.5 were used in examining LOCHWISP's searching behaviour.

(c) Searching with overlapping catchments

Initially, the greatest uncertainty about the new algorithm was whether it would work with overlapping catchments. With Thiessen catchments it is easy to see that, since the allocation stage can only shorten the distance between a demand point and the centre to which it is assigned, it must improve the value of the objective function obtained at the end of the location stage. However, when the allocation of points of demand to centres is defined by a spatial interaction model such as 7.1, it is not clear at the outset whether the allocation stage will necessarily improve whichever objective is being sought by the location stage.

Before discussing how the two stages work together with overlapping catchments, it is helpful to discuss some of the implications of the model used to define these catchments. As noted earlier, the catchments are defined by the parameter b_s , used as the exponent of distance in the spatial interaction model forming the basis of subroutine SHARE. If b_s has a fairly high value, such as 0.5 then a relatively high proportion of the population in any grid cell will be assigned to facilities which are relatively close, thus forming compact catchments. If b_s is given a lower value, say 0.25, the fraction of the cell's population assigned to relatively more distant facilities increases and all facilities have wider catchments with greater areas of overlap between them. Consequently, a significant proportion of the population of any cell could then be assigned to each of the three or four closest centres.

The value of b_s is not related to the spatial elasticity of demand as defined by the parameter, b . For instance a value

of 0.5 for b_s could equally well be associated with relatively inelastic demand such as $b = .125$ or strongly elastic demand such as $b = .50$. Since the catchments of most pools in Edinburgh tend to be fairly compact (Currie, 1977), it seems appropriate to use larger values for b_s than for b . In exploring the patterns of search for overlapping catchments, values of 0.25 and 0.50 have therefore been used for b_s , in combination with each of the three standard values of b used earlier, potentially yielding six combinations of catchment size and elasticity. The combination of $b = .5$ and $b_s = .25$ has not been used, however, since this would imply a lower value for b_s than for b , which leaves only five combinations. It is worth emphasising that the allocation stage merely involves distributing population and therefore, implicitly, a constant level of potential demand between surrounding centres. Hence the overall level of use in the system cannot deteriorate during this stage of the algorithm. Since each value of b_s produces a different size of catchments, naturally each yields a different value of cost of travel, AT , for the system. Thus when demand is inelastic there is no longer a single value of cost of travel associated with each problem as in the case with Thiessen catchments.

An example may clarify how the two stages work together. Consider the problem of locating a second facility to serve Edinburgh given that one is already fixed in position, just to the east of the centre of the city. Suppose the second facility is initially sited slightly to the west of the centre.

Initially catchments will be defined around each facility during the first execution of the allocation stage. Each centre then draws

population from a catchment rather like the demand cones of Chapter 3 in shape in that the closer a demand point is to facility j the larger the share of its population assigned to j . When the location stage is carried out within the catchment of the second facility it will be moved to a point further west, thereby reducing the travel cost and/or increasing the levels of use and NBU purely within this centre's own catchment. Of course, within this catchment some points will be closer to the new position and some will be further away but, since the method of iterative search has been shown to work for all three objectives, the net effect must be to improve the objective concerned within the catchment as a whole.

Since the other facility is fixed the values of the various objectives within its catchment will remain the same. Thus, if the objective being sought is evaluated over both catchments at the end of the location stage and before re-allocation, some improvement must have occurred. This value of the objective function can be taken as a notional bench mark before the re-allocation stage. It is worth emphasizing that the catchments at this point are still based on the allocations defined from the initial locations and do not take account of the fact that one centre has been moved to a new location.

Consider now any demand point i from which a certain portion of the population was assigned to the first and a certain position to the second facility according to their initial locations. At the end of the location stage the second facility may be closer to i or further away than before. When the allocation stage is carried out for the second time, if this facility is now closer,

more of i 's population will be assigned to it; if the facility is further away, less of i 's population will be assigned there. In either case, the cost of travel which originates from i must be lower and the levels of use and NBU must be higher than they were with the locations and pattern of catchments existing at the end of location stage. This must be true of all points of demand. Thus, in this example the allocation stage will always improve the situation obtaining at the end of the location stage before the catchments have been adjusted to take account of the new location.

If we consider a problem where several moveable centres are to be located, it is clear that the location stage will always improve the objective function within each separate catchment. Furthermore, when the locations of facilities change relative to a particular demand point, the allocation stage will always allocate more population to a facility which has moved relatively closer and less to one which has moved further away in relative terms. Thus the allocation stage must bring an improvement in the objective concerned for the whole system of centres. Thus the location and allocation stages must inevitably work together in harmony to produce convergent results for each of the three objectives which LOCHWISP can treat.

The inevitability of convergence was not foreseen and was only realised from the results of trials involving overlapping catchments. The spatial interaction model was originally incorporated in LOCHWISP in an attempt to ensure that a facility would be repelled by other centres when searching with overlapping catchments. It was initially thought that, with this modification,

LOCHWISP might produce convergent results with some problems but not others. It was the fact that convergence always occurred which prompted a closer examination of its behaviour and yielded the rationale just outlined.

When catchments overlap the searching behaviour of LOCHWISP is somewhat different from that already described for Thiessen catchments. These differences can perhaps be characterised most easily by comparing the results for problems concerned with maximizing use. Since overlapping catchments involve the use of an extra parameter, b_s , first of all it is necessary to say what influence b_s has on the results. The higher value of b_s , .50, naturally produces more compact catchments and a stronger tendency for facilities to be repelled by each other during the process of search. As a result, when $b = .125$ and $b_s = .50$, the final locations and the lines of search are very similar to those produced by using Thiessen catchments with the same value of b (Fig. 7.8). In fact the latter results were very similar to those already obtained in minimizing the cost of travel within Thiessen catchments, which was the most dispersed set of locations found in the previous series of tests. By contrast, when catchments overlap with $b = .125$ and $b_s = .25$ the final locations are much less dispersed and a loose cluster of facilities is formed on the eastern side of Edinburgh (Fig. 7.9).

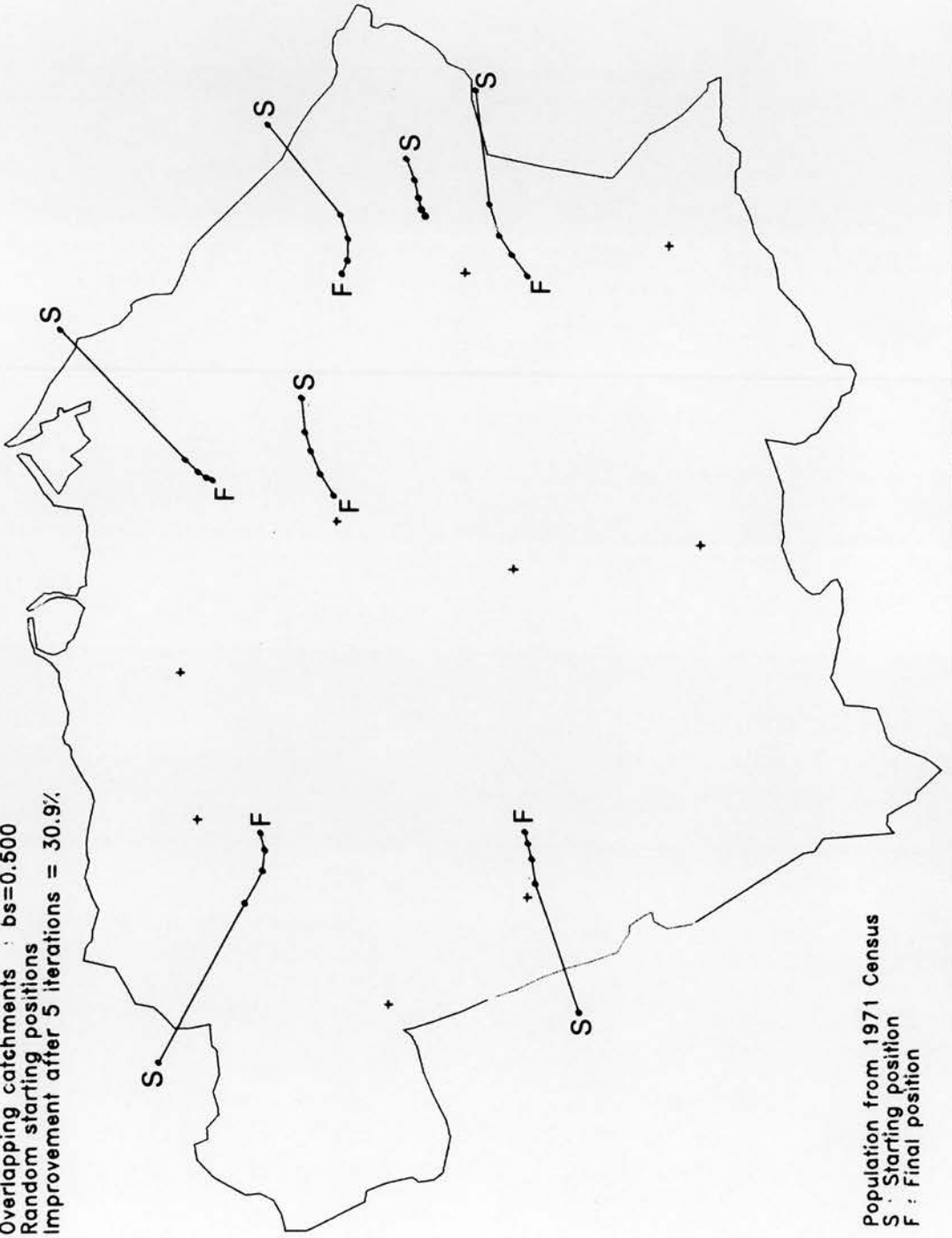
Though the different values of b_s yield quite different results when $b = .125$, it should be noted that their results are much more similar when $b = .25$, those for $b_s = .50$ showing only a slightly stronger tendency to disperse from their initial locations, (Fig. 7.10). Compared to both sets of results for

Figure 7.8 Lines of search for location of 7 centres in Edinburgh
to maximize use with $b=0.125$

Overlapping catchments : $bs=0.500$

Random starting positions

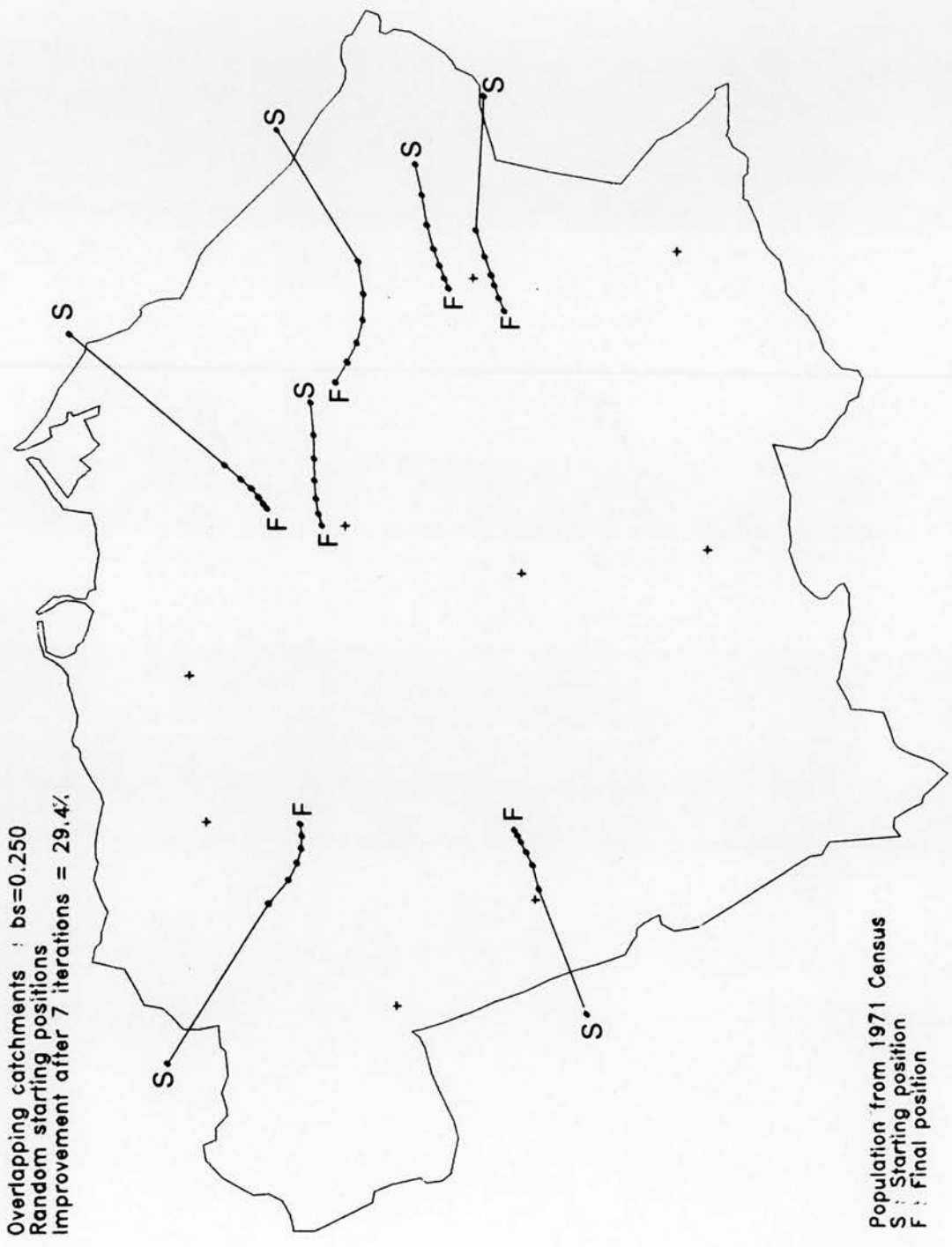
Improvement after 5 iterations = 30.9%



Population from 1971 Census
S : Starting position
F : Final position

Figure 7.9 : Lines of search for location of 7 centres in Edinburgh
to maximize use with $b=0.125$

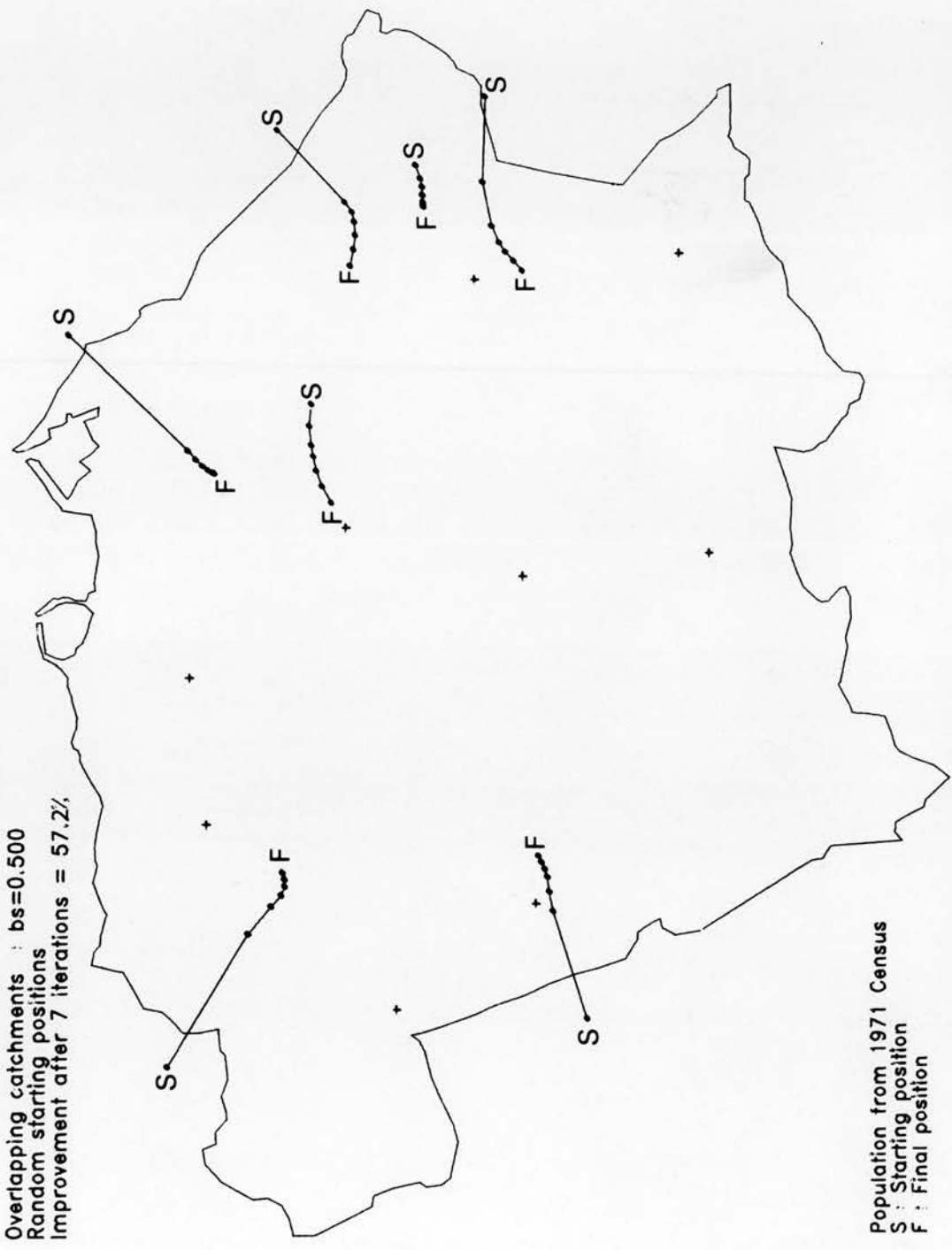
Overlapping catchments : $bs=0.250$
Random starting positions
Improvement after 7 iterations = 29.4%



Population from 1971 Census
S : Starting position
F : Final position

Figure 7.10: Lines of search for location of 7 centres in Edinburgh
to maximize use with $b=0.250$

Overlapping catchments : $bs=0.500$
Random starting positions
Improvement after 7 iterations = 57.2%



Population from 1971 Census
S : Starting position
F : Final position

$b = .125$, those for $b = .25$ show a somewhat stronger tendency to cluster on the east side of Edinburgh, mainly because they fail to move very far from their initial positions irrespective of the value of b_s . Thus, with Thiessen catchments, a higher degree of elasticity can more easily yield a more clustered pattern of locations, partly because the search stops at nearby local optima more frequently.

The results already shown for $b = .125$ and $b_s = .25$ highlight particularly well the general difference between the results for overlapping and for Thiessen catchments. The greater attraction of the city centre is evident from Figure 7.9 on which the lines of search almost all trace paths towards the city centre, as they do to a lesser extent on Figure 7.10. The greater attraction of the city centre reflects the need to find good locations for serving wider tributary areas than Thiessen catchments would allow, so this attraction is naturally felt more strongly when $b_s = .25$. Even with the relatively dispersed solution given by $b = .125$ and $b_s = .50$, the final locations tend to be displaced somewhat towards the city centre compared to their Thiessen counterparts. A similar difference occurs with $b = .5$ and $b_s = .50$ in that the facility which moved to a position in the south east of the city with Thiessen catchments is moved instead only a short distance towards the centre of the city. This movement towards the centre, observed in almost all instances is facilitated by the weaker repulsion of the centres for each other when catchments overlap. Together, the weaker mutual repulsion and the attraction of the centre produce the cluster of facilities frequently observed to the east of the centre. The fact that this often leaves the

southern part of the city unserved suggests that overlapping catchments are more likely to yield only local optima.

Another general difference is that, although the lines of search for several overlapping problems travel quite far from their initial positions (e.g., Fig. 7.8), on the whole the facilities move less far than with Thiessen catchments, albeit often in a similar direction. This also partly explains why local optima appear to be more frequent. As before, it is noticeable that the first step in the line of search is always relatively long, though perhaps not quite so long as with Thiessen catchments.

The pattern of results obtained from maximizing NBU was on the whole rather similar to that for maximizing use with the tendency to stick at local optima even more marked. When travel cost is minimized the results for overlapping catchments are less different from the Thiessen equivalent than when use is maximized. With $b_s = .50$ some of the facilities are displaced a short distance towards the centre; with $b_s = .25$ this displacement is very much greater.

The main conclusion from the preceding discussion is that, in theory, an algorithm for locating central facilities with overlapping catchment areas must always converge, provided it uses the same basic principles as LOCHWISP. The results of the trials carried out confirm this conclusion in practice.

In modifying LOCHWISP to treat overlapping catchments the main difficulty lay in introducing a spatial interaction model into the program with as little alteration as possible to the program's existing structure. Though this required care,

relatively few practical difficulties were encountered after the basic ideas for doing so had eventually been formulated clearly. The solution of overlapping problems with 7 centres needed roughly three or four times as much time for computation as the equivalent Thiessen problem. Because of its evolution from a program mainly designed for Thiessen problems, however, LOCHWISP is somewhat inefficient for certain forms of the overlapping problem and could easily be made more efficient.

The introduction of a spatial interaction model in fact gave scope for allowing centres to vary in attraction. Unfortunately the latter development came too late in the work to be fully used in analysing the locations of swimming pools. Nevertheless, after the chapters on swimming pools had been completed it was used to re-examine some of their conclusions.

Conclusion

The main conclusion of this chapter is that it is possible to find sets of locations which maximize demand or net benefit to users by using a searching method guided by the gradient of those functions and that this method can be applied successfully to overlapping as well as to Thiessen catchments. Apart from the fact that local optima may be more frequent when demand is very elastic and catchments overlap, a location/allocation algorithm based on these principles works almost as efficiently and satisfactorily as those already in existence for minimizing aggregate travel cost.

With Thiessen catchments the main factor influencing the locations obtained is the assumed degree of elasticity. Relatively inelastic or moderately elastic values of b all yield very similar results. Though these were sometimes quite different from the solutions for strongly elastic values of b , often the strongly elastic and fairly inelastic solutions were not far from each other. Basically the various goals represent different ways of measuring accessibility. If Thiessen catchments are used it seems that in siting service centres within a city such as Edinburgh whichever of the three measures is chosen may not make as much difference to the final result as might have been expected.

With overlapping catchments the degree of elasticity and the extent of overlap both strongly influence the outcome and facilities generally tend to be more attracted by the city centre than with Thiessen catchments. The results also suggest that there is then more difference between the solutions for different forms of the problem, as defined by the various goals and their associated degrees of elasticity and of overlap. The way in which accessibility is measured is therefore more important in this context.

One of the most significant findings of the present work is that an alternating type of algorithm which uses a simple spatial interaction model to define overlapping catchments must logically produce convergent results. Moreover, since it is not particularly difficult to write the program for such an algorithm, there is much less reason for Thiessen catchments to be used in future work of this type in situations where they may not be very appropriate.

SECTION IV

The location of swimming pools in Edinburgh

CHAPTER 8

The location of the Commonwealth Pool

Introduction

Having established that LOCHWISP works, we can now use it to examine the location of swimming pools in Edinburgh. Accordingly our main aim in this section is to answer three questions:

- (a) What factors influence decisions made by the local council to locate new pools at certain points rather than others?
- (b) How satisfactory is the present distribution of pools in Edinburgh in terms of spatial efficiency and equity?
- (c) Are the criteria incorporated in LOCHWISP a significant improvement on those used at present by the local authority in choosing locations?

Since comprehensive information on the distribution of population after 1971 is not available, the 1971 Census of Population has to be used to define the distribution of demand. Fortunately for our research the only two public indoor pools built in Edinburgh since 1900 were opened in 1970 and 1978, so that data for 1971 are very relevant to the former. A 1971 data base is obviously rather unsatisfactory for examining a facility whose construction started in 1975 and finished in 1978 but there was no alternative.

The present chapter will be concerned with the decision to locate the first of these pools, the Commonwealth Pool; the choice of Wester Hailes for the location of the new district

pool opened in 1978 will be examined in the next chapter.

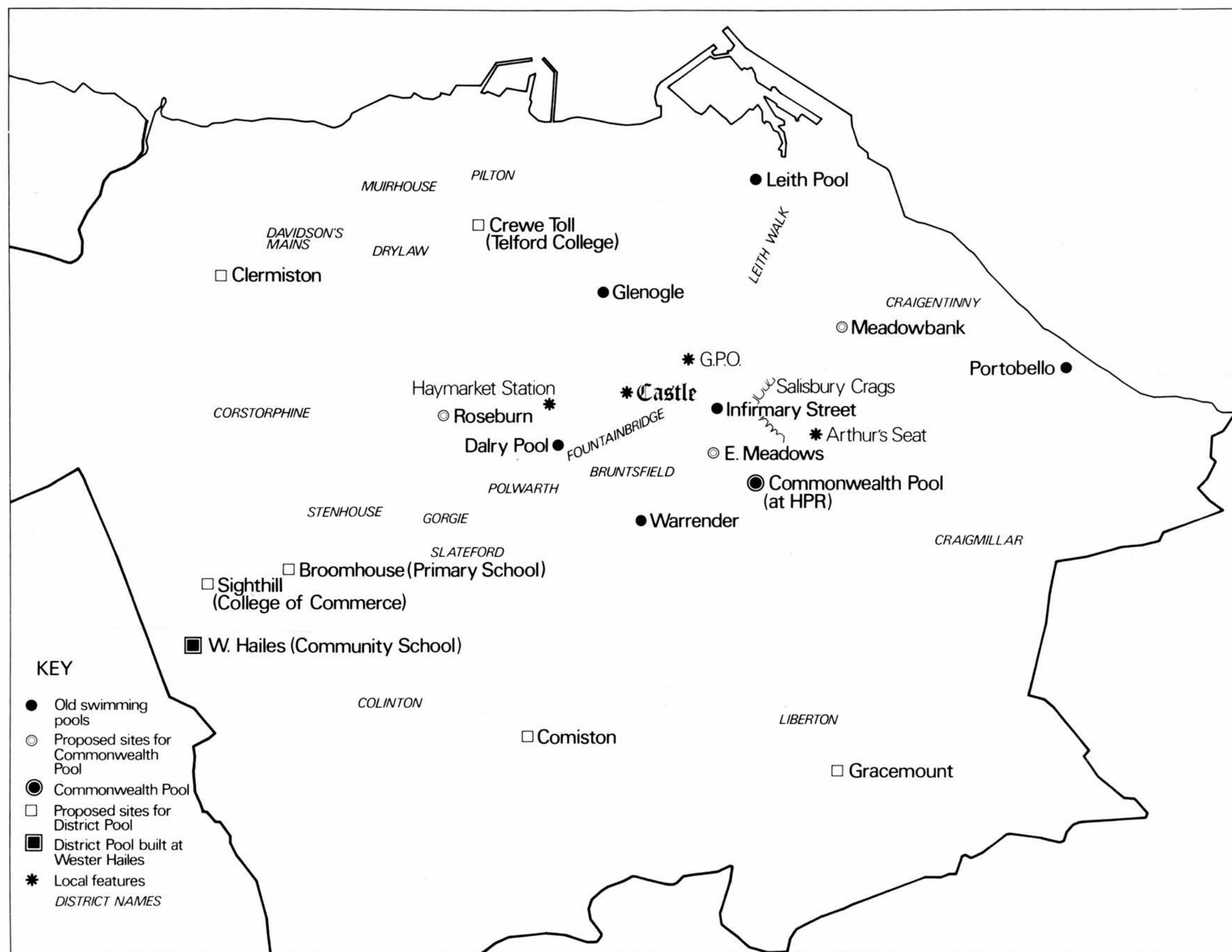
The location of pools in Edinburgh

The most striking feature of the distribution of the six old public baths in existence before the Commonwealth Pool was built is that most of them are fairly close to the centre, with the notable exception of the indoor pool at Portobello on the eastern side of the city (Fig. 8.1). The construction of the Commonwealth Pool did not in fact change this pattern since it was also located fairly near the centre. Thus, as late as 1977, there were no pools in any of the western suburbs of the city.

An obvious question is whether such an arrangement is efficient. This question can be answered by locating six pools to optimize various goals, using in turn Thiessen and overlapping catchments. Using LOCHWISP, a set of locations can be found to which the cost of travel within Thiessen catchments is 36.2% lower and the maximum travel distance is 37% lower than to the six old pools. Furthermore, if we find the set of six locations which appears to maximize use with $b = .25$, its level of use is 24.3% higher than that for the six old pools. Similarly, a set of locations, not necessarily optimal, can be found whose value for NBU (with $v = 5.0$ and $b = .25$) is 31.2% higher than it is for the latter. Similar levels of improvement can be found for other levels of b and v .

If overlapping catchments are used, when $b = .25$ and $b_s = .50$ an improvement of 19.6% in total use can be obtained. In addition, with $b = .125$ and $b_s = .50$, one run of LOCHWISP produced an improvement of 16.9% in use. These results are not based on exhaustive trials to find the global optima, so greater levels of

FIGURE 8.1 REFERENCE MAP OF EDINBURGH



improvement could probably be attained. In each of the cases just described the locations of the old pools were used as starting positions for search. During each search several pools were usually moved westward; those originally sited at Dalry and Glenogle were often moved more than 2 or 3 km. further west.

Thus the existing distribution of pools in the city in 1968 was rather inefficient and rather inequitable in spatial terms. It is therefore important to understand why this situation developed.

History of swimming baths in Edinburgh

All the pools in existence in 1968 were built in the 1880s or 1890s and were therefore located to serve the relatively compact city of the turn of the century. If we consider them in relation to the city as it was then they appear to cover the urban area rather efficiently. Glenogle Road covered the north/northwest, Dalry the west, Warrender the south and Infirmary Street the central area (Fig. 8.1). The remaining pools (at Leith and Portobello) covered what were then two of the main outlying parts of the built up area to the north and east respectively. Since their distribution ensured that most citizens had good access to at least one pool, these locations may well have been the result of a systematic spatial plan.

In the interwar and postwar period the city expanded rapidly in all directions but particularly south and west. As a result the extensive new residential areas in the south and west enjoyed rather poor access to the existing baths. The fact that the first indoor pool built by the city after 1914, the Royal Commonwealth Pool, did not open till 1969 raises questions about why the city council was so slow to improve access to such a popular and growing

form of recreation. Of course, the primary reason for this unsatisfactory distribution in 1968 was the city's rather eccentric growth in the interwar and postwar periods, but the council's apparent slowness to build new pools may also have been a factor.

Though an explanation for the delay in constructing new pools lies outside the scope of the present work, it is relevant here to ask why the new pool was built fairly near the centre, only $1\frac{1}{4}$ km and $1\frac{1}{2}$ km from the existing pools at Infirmary Street and Warrender respectively, rather than on the western or southern sides of the city. To answer this question we have to consider the history of this particular decision and examine the deliberations and plans of the bodies involved in making it.

The decision to locate the Commonwealth Pool

During the period under discussion the committee of the city council (Edinburgh Corporation as it was till 1974) responsible for bringing forward plans concerning swimming pools was the Civic Amenities Committee (hereafter abbreviated to CAC). One of the sub-committees of CAC was the Baths and Laundries Sub-Committee (BLSC). Plans and reports concerned with pools were normally first brought to BLSC for detailed discussion; recommendations were then passed on to CAC.

The construction of a new pool was first mooted in BLSC in December 1958 when a report by the City Architect suggested possible sites for a 'new central pool' in connection with the possibility of the Commonwealth Games being held in Edinburgh at a future date but the decision to build the pool on its present site was not finally taken until early in 1965. The minutes of BLSC

and CAC between 1958 and 1965 and some relevant reports kindly made available by City of Edinburgh District Council (CEDC) throw much light on this decision and have therefore been examined with a view to answering the following questions:

- (a) what sites were seriously considered?
- (b) why was the present location chosen?
- (c) what criteria did CAC and BLSC use to assess their relative locational advantages, in particular how much weight was given to accessibility to users and how was it measured?

From the beginning the new pool was always referred to as 'a central pool'. The idea that a major pool could be sited on the western side of the city to meet local needs on an everyday basis yet still satisfy the more specialised needs of a wider region for racing and diving was never mentioned in the ensuing discussion, as far as can be judged from the minutes. A site at Roseburn, about 2.5 km to the west of the centre, was a very strong candidate early in the deliberations but one of its main virtues was seen to be its proximity to the centre and to regional transport facilities such as Haymarket Station. It was never recommended as a means of serving the needs of the western parts of the city.

The various minutes and reports available give little direct indication of why it was taken for granted that a central location was essential. A number of factors, however, are implicit in the assessments made of possible sites. First, it was stressed that the pool was to have a regional function and would therefore be expected to attract visitors from other parts of the Lothians and Fife. A central location near railway and bus termini and well-

sited with respect to the new road system then planned for the city (though partially abandoned at a later date) may have seemed essential for this regional function. It was also felt by the City Architect that a visually prominent site was desirable, partly to attract users but also because the civic and symbolic importance of the building warranted it. This requirement was much more likely to be satisfied at a site near the city centre which would be more visible to passing traffic.

Despite their preoccupation with the new central pool the committee and sub-committee were aware that there was a gap in provision on the southern and western sides of the city. In October 1960 BLSC agreed in principle to make arrangements for the construction of four district pools to fill this gap. This decision was affirmed in principle on numerous occasions thereafter, the four sites usually mentioned being at Comiston, Crewe Toll, Sighthill and Clermiston or Liberton (Fig. 8.1). Because plans were afoot for the western and southern areas, presumably CAC and BLSC felt they could ignore the local needs of those areas in siting the new pool built to Olympic standards for the Commonwealth Games.

However, in the twenty years since 1960 only one district pool has actually been built, viz., that at Wester Hailes. It may well be worth asking how strong the political will to build the district pools really was in the first place. A sceptic might be forgiven for thinking that if Edinburgh had not been willing to invest in a single indoor pool between 1900 and 1960 it was unlikely to tax the ratepayers with five pools in one decade, one of them to Olympic standards, however much the sport

had grown in popularity. Nevertheless, the existence of the plans for the district pools probably helps to explain why it was acceptable to assume that the Commonwealth Pool should be located near the centre.

Between December 1958 and January 1965 at least eleven sites were considered as possible locations for the new pool, none more than a mile and a half from the city centre (defined as the General Post Office in Waterloo Place). The present site only came into the reckoning rather late, in January 1965. Before that, decisions had been made in principle by CAC to build the pool first at Roseburn Park, then at Meadowbank; later still a site at the eastern end of the Meadows was very seriously considered.

On the 20th July 1960 a joint meeting of the Planning and Civic Amenities Committee voted decisively for Roseburn as opposed to Meadowbank and a site at Hope Crescent. On the 27th July a report in 'The Scotsman' stated that 900 local residents had signed a petition opposing the choice of Roseburn. The main reasons for opposition were given as the loss of amenity and of playing fields in Roseburn Public Park, the extra traffic congestion in the area and the objection in principle to the alteration in the city's development plan which the proposal entailed.

The first of these reasons was probably the main one. For instance, two letters published in 'The Scotsman' on July 25th 1960 were very much concerned with the amenity issues. It should, however, be noted that the pool only required part of the park, about one third approximately. It is possible that the strength of the objection partly derived from the local residents imagining that the pool would somehow resemble the existing Victorian baths in appearance

and clientele and would not therefore add to the quality of a rather pleasant residential area with relatively few young people, but this is only speculation. Whatever the real reason behind the objections, whatever the rights and wrongs of the issue, the protest was effective. After this date the Roseburn site was more or less dropped from consideration.

On 25th January 1962 a Special Joint Committee of CAC and the Planning Committee expressed a view that the new baths should be at Meadowbank rather than Roseburn. No action had been taken on this view by 28th July 1964 when a Joint Sub-Committee of the Lord Provost's Committee and CAC reaffirmed this recommendation by voting eight-four in favour of Meadowbank as opposed to a site in The Meadows. A meeting of the full city council two days later, however, resolved that more consideration be given to The Meadows site.

On 19th January 1965 it was reported to the Joint Sub-Committee that the owners of the ground at the junction of Dalkeith Road and Holyrood Park Road, (then known as Park Road) had said that they would be willing to sell this ground to the council for the construction of a pool. It seems to have been agreed quite rapidly that this site was far superior to the previous ones; with almost no opposition it was quickly decided that the pool should be built there instead of Meadowbank or The Meadows. Most of this land was already zoned for recreation so there was no difficulty over planning permission.

It is interesting to ask why agreement on the site at Holyrood Park Road (hereafter abbreviated to HPR) should have been reached so quickly. First, the physical setting with Salisbury Crags and Arthur's Seat in the immediate background is

magnificent, more than meeting the requirement for a prominent site. Second, there were few people living next to the site so there was likely to be little opposition, whereas the site at The Meadows would inevitably have been opposed on amenity grounds. Third, since the site was available, work could start on it quite soon: with the Commonwealth Games for 1970 having recently been given to Edinburgh, speed was becoming rather important. Fourth, it was conveniently close to Meadowbank where the athletics events of the Games were to be held and it was immediately adjacent to the Pollock Halls of Residence where the athletes would probably live during the Games.

Having discussed the history of the decision to locate the pool at HPR, we can also use these minutes and associated reports to gain an impression of how the factor of accessibility was treated and measured during the discussions of CAC and BLSC and to assess how much weight it received. Apart from distance from the city centre, accessibility was mainly measured by the number of bus routes passing near a site and by what was termed the 'ambient population' of a site. The number of bus routes is a very useful index of a location's access to the rest of the city and should certainly be given for any site under consideration, especially for a central pool. Nevertheless, if sites with many passing bus routes were always given preference, no pools would ever be built in the suburban areas. As far as 'ambient population' is concerned this was never defined very explicitly. It seems to have meant the population within a radius of a mile or a mile and a half, rather than the discrete catchment defined for instance by the Thiessen polygon around the point or by another

method which takes account of the location of other pools. Defined in the former way, 'ambient population' may be quite a useful measure, but two pools quite close to each other could then both be said to have large ambient populations, ignoring the fact that in reality this large population would have to be split between them. Thus the site at HPR was said to have about 40,000 within one mile but of course the radius also encompassed the baths at Infirmary Street and Warrender.

In fairness it should be said that reports by the City Architect usually noted how near existing pools were at any site, but little attempt was made to assess what impact the new pool would have on existing pools nearby and vice versa. In fact, apart from the exception just mentioned, ambient population was never actually quantified as far as can be judged from minutes and reports. For instance, a report in 1960 simply states that for Roseburn "there is a substantial population in the districts of Stenhouse, Corstorphine and Clermiston."

Though access to users was one of the criteria taken into account, it is difficult to gauge its relative importance from the records available. It may be significant that ambient population and relationship to existing pools usually came after availability of site, visual impact, bus transport and parking space on the list of criteria used, in reports presented to CAC and BLSC, but the order of these factors could equally well be a matter of chance. One point, however, does emerge from a study of these records. Unless locational models are fundamentally wrong, location would inevitably be a very important, perhaps the most important, factor influencing who would use the new pool and how

often. Yet this point never appears to have been stated explicitly, let alone stressed as it ought to have been.

Admittedly there were difficulties. In the early 1960s it was more difficult to obtain accurate data on population at a sufficiently detailed spatial scale to allow potential catchment population to be estimated. The fact that it was much less easy to do so may partly explain why access to users was not examined very systematically, but it cannot really explain why this factor was inadequately stressed. And had it been considered desirable to compute ambient population more accurately, this could have been done by using school catchments and the electoral register.

Analysis of potential locations by LOCHWISP

If it is accepted that the choice of location for the new pool is the major factor determining its impact on access to swimming facilities in the city as a whole, and its own pattern of use, we can employ algorithms such as LOCHWISP to examine this choice of location. One problem in doing so is that NORLOC assumes and the original version of LOCHWISP assumed that centres are uniform in character, that they provide essentially the same service and are therefore equally attractive to users. This seems a reasonably acceptable assumption to make about the six Victorian baths in the city which are similar in size as well as design and appearance. The Commonwealth Pool, however, has special facilities for diving and for toddlers; its main pool is roughly 4.25 times the size of all the older pools, except that at Infirmary Street. It also has a cafeteria and good facilities for spectators and competitors. For these reasons it must be viewed as much more attractive than the

six older pools. With respect to these specialised services it therefore serves a wider area than the older pools, particularly in summer as the Scottish Sports Council (1979) study shows.

To do justice to the Commonwealth Pool we therefore require a location model which takes account of attractiveness. As noted earlier, most of the analysis carried out in this chapter was done first with a version of LOCHWISP which did not allow the attractiveness of facilities to be considered, partly because the spatial interaction model allowing catchments to overlap was only successfully incorporated in the program at a relatively late stage in the work. Although it later became possible to include attractiveness in this model, at first it had been decided not to do so partly because of the limited time available for completion of the work, and partly for technical reasons which will be discussed presently. However, after both these chapters had initially been completed, LOCHWISP was modified to include attractiveness. The spatial interaction model used to allocate the potential demand at i , p_i , to a facility at j then had the form

$$T_{ij} = p_i w_j e^{-b_s d_{ij}} / \sum_{j=1}^m w_j e^{-b_s d_{ij}} \quad (8.1)$$

where w_j is the attractiveness of centre j .

Some results were recomputed on this basis and will be presented later where relevant.

Ideally, a spatial interaction model incorporating attraction ought to be calibrated to fit the existing pattern of movement to all seven pools. Having obtained a measure of attractiveness for the Commonwealth Pool in this way, we could then apply it to the spatial interaction model just described. This approach requires

a large amount of data to be collected and in the time available a simpler method was needed.

Attractiveness was measured by the area of the pool itself. Five of the six old pools measure 22.85 m by 10.66 m and have been given a value of 1.00 for attractiveness; with a slightly larger surface area of water the pool at Infirmary Street has a value of 1.22. In measuring the attractiveness of the Commonwealth Pool, the small pool for toddlers has been included, giving a value of 5.66. This probably underestimates the attraction of the latter because it takes no account of its diving pool or the other services it provides for users and spectators.

The earlier decision to omit attraction can, nevertheless, be defended on several grounds. First, the Commonwealth Pool does serve the local area around it. Studies of its catchment area show that, especially in winter, there is a tributary area around it which is much broader but not unlike that around the other pools (Currie, 1977; Scottish Sports Council, 1979). We can therefore argue that, in omitting attraction, we assess the Commonwealth Pool as a local pool, which is part of its role.

A much more important point is that, when used without attraction, we can view LOCHWISP purely as a way of assessing the accessibility characteristics of locations in space regardless of whether these locations are to be used for pools, libraries or surgeries, large or small. In this form, LOCHWISP provides a means of finding sets of accessible locations which may then be exploited by facilities of high or low attraction. Moreover, it is legitimate to argue that attractiveness, as opposed to accessibility, is basically a property of the facility, not the location. A given type of facility will therefore have the same

attraction at any location within the city. Whether its attraction is great or small, to maximize use we only have to pick the most accessible location, taking into account the locations of existing centres.

To some extent this argument allows us to ignore the attractiveness of the facility in picking a location. In a sense it is equivalent to arguing that attractiveness and accessibility are mathematically independent and this is largely the case in the model described in 8.1.

To assess the present location for the new pool we can first of all compare it with the alternative sites at Roseburn and Meadowbank using Thiessen catchments. The Meadows site has not been included in this comparison because it is fairly close to the site at HPR and was invariably found to have similar characteristics in terms of accessibility. To make comparison easier the values for each of the other sites have been expressed as a percentage of the values for the location at HPR, as have the equivalent values with only the six old pools in existence (Table 8.1).

When examining these values it is worth recalling that the index of travel cost assumes inelastic demand, whereas the index of use, TRIPS, assumes a significant decline in frequency of use within the catchment of that centre. As noted earlier, to determine how spatially elastic the demand for swimming is would involve a separate research project, but some preliminary evidence from a study by Currie (1977) suggests it is moderately elastic. In an empirical sense travel cost is not therefore the most valid measure of accessibility for this particular service. However, as noted earlier, travel cost is a more equitable measure than TRIPS

Table 8.1

Indices of accessibility for all pools with a new pool at various sites, expressed relative to the values for Holyrood Park Road, assuming Thiessen catchments

<u>Index</u>	<u>Six old pools</u>	<u>New pool at H.P.R.</u>	<u>New pool at Roseburn</u>	<u>New pool at Meadowbank</u>	<u>New pool at 'efficient' location (8.1; 11.0)</u>
<u>Travel cost</u> (AT)	102.7	100	90.5	99.5	80.4
<u>Maximum distance travelled</u>	100	100	81.1	100	69.5
<u>Use</u> (TRIPS)					
(1) $b = .125$	98.4	100	103.1	100.4	109.1
(2) $b = .25$	97.6	100	104.9	102.4	112.8
(3) $b = .5$	95.6	100	102.8	103.7	116.5
<u>NBU</u>					
$v = 5.0$					
$b = .25$	96.1	100	103.8	108.5	114.1
<u>Population covered</u>					
$s = 5.0$	96.6	100	105.1	96.8	120.2

and is of interest because it combines elements of efficiency and equity. An earlier study of the location of swimming pools in Lothian Region by Cargill and Hodgart (1978) unfortunately failed to draw attention to this limitation of travel cost.

All the values in Table 8.1 are total figures for the whole system, i.e. all catchments added together, rather than for the new locations themselves. On all the measures of accessibility shown for Thiessen catchments the site at Roseburn brings an improvement which is superior to that for the site at HPR. On all the measures except one, NBU, the former site is also superior to Meadowbank. On most measures the Meadowbank site is actually slightly better than the site at HPR. This is mainly because it fills the centre of the interstice between the pools in Leith, Infirmary Street and Portobello, whereas the site at HPR is quite near the two old pools at Infirmary Street and Warrender.

In comparing results based on overlapping catchments it should be noted that the maximum distance travelled is merely the distance to the grid cell which happens to be furthest from the site under consideration. Since it is therefore meaningless as a measure of equity of access, it is not given for overlapping catchments. The population within the covering radius of facilities is also omitted for overlapping catchments because its pattern of results was almost invariably similar to those for TRIPS.

As with Thiessen catchments, the site at Roseburn produces significantly lower travel costs and higher levels of use in the system than the site at HPR for all the values of b and b_s (Table 8.2).
used with overlapping catchments/ When catchments overlap, each site can potentially collect demand from the whole city and both

TABLE 8.2

Cost of travel and level of use for all pools with
a new pool at various sites, expressed relative to
the values for HPR, assuming overlapping catchments

	<u>Six old pools</u>	<u>New pool at HPR</u>	<u>New pool at Roseburn</u>	<u>New pool at Meadowbank</u>	<u>New pool at 'efficient' location (8.1;11.0)</u>
<u>Cost of Travel</u>					
(1) $b_s = .25$	99.4	100	95.1	100.4	90.0
$b_s = .50$	100.2	100	92.4	99.6	83.6
<u>Use (TRIPS)</u>					
<u>$b = .125$</u>					
(1) $b_s = .25$	100.6	100	103.1	100.3	107.5
(2) $b_s = .50$	99.9	100	103.8	100.7	109.5
<u>$b = .25$</u>					
(1) $b_s = .25$	101.2	100	104.5	101.2	113.1
(2) $b_s = .50$	99.9	100	105.8	101.8	115.3
<u>$b = .50$</u>					
(1) $b_s = .50$	99.9	100	104.8	103.9	119.7

locations are relatively central in the city. Thus the differences between the sites on all indices of accessibility tend to be less great than with discrete catchments. In general, the greater the degree of overlap the more the relative advantages of one site tend to be blurred compared to a less accessible site. As a corollary, the numerical differences between good and bad sites tend to be maximized when Thiessen catchments are used. The fact that the site at Roseburn again produces higher levels of demand is of course mainly because it is nearer the poorly served western areas of the city. Comparisons were also made between the three sites in terms of NBU with both types of catchments but the results were very similar to those for use and are therefore not shown.

One puzzling feature of Table 8.2 is the fact that with $b_s = .50$ travel cost to the six old pools is very slightly lower than it is after the addition of a pool at the site in HPR. The extra pool must reduce the travel costs of nearby users, so it can be argued that travel costs over the whole system must be reduced. However, since this site is near several existing pools and some users from peripheral locations on the west side of the city will be assigned there by the spatial interaction model, a small number of people will be allocated to it who had previously been assigned to nearer pools at Glenogle or Dalry, thereby increasing travel costs by more, apparently, than the reduction just mentioned. The fact that a few journeys are longer than before can also tend to reduce demand, which explains why the overall level of use associated with the site at HPR is very slightly lower than with the six old pools for some values of b and b_s .

These rather curious results only occur when pools are added at what are relatively poor locations in terms of the model, i.e.

near existing facilities or rather far from any of the city's population. A short series of tests using locations distant from the city confirmed this point. Clearly this feature of the model is unsatisfactory and a means of reformulating the model so that it does not occur would be desirable.

When attractiveness is taken into account, the results are similar in that Roseburn is again the best of the three sites and Meadowbank is slightly better than the site at HPR on several measures (Table 8.3). However, the differences are often much larger than before, with Roseburn more than 9% better than HPR on five of the seven indices. When the pool placed on the western side has its attractiveness increased it draws a greater share of the very large population in the western half of the city, thereby shortening travel distances and increasing the overall level of use. On the other hand, when attractiveness is included the site at HPR naturally draws more users than before, but many of these have to travel long distances because there are other pools nearby and so the overall impact on cost of travel and use is less.

Thus the comparison of the three sites using overlapping catchments reaches similar conclusions to that based on Thiessen catchments. On all measures of efficiency, the site at Roseburn is superior to the sites at Holyrood Park Road and Meadowbank particularly when attraction is taken into account. Since the value of TRIPS is consistently higher at Roseburn, the city's total revenue from admissions to pools would probably be greater if that site had been chosen in preference to that at HPR. If Thiessen catchments are assumed, Roseburn also yields the best value for equity of access.

TABLE 8.3

Cost of travel and level of use for all pools with
a new pool at various sites, expressed relative to
the values for HPR, using overlapping catchments
and taking attractiveness into account

	<u>Six old pools</u>	<u>New pool at HPR</u>	<u>New pool at Roseburn</u>	<u>New pool at Meadowbank</u>
<u>Cost of Travel</u>				
(1) $b_s = .25$	95.0	100	90.7	102.2
(2) $b_s = .50$	95.1	100	87.5	99.3
<u>Use (TRIPS)</u>				
<u>$b = .125$</u>				
(1) $b_s = .25$	104.5	100	106.5	99.9
(2) $b_s = .50$	103.4	100	106.9	101.2
<u>$b = .25$</u>				
(1) $b_s = .25$	109.0	100	109.9	102.0
(2) $b_s = .50$	106.2	100	109.9	103.0
<u>$b = .50$</u>				
(1) $b_s = .50$	110.2	100	110.3	106.2

Effect of the Commonwealth Pool on catchments and use of older pools

(a) Predicted impact

Though never mentioned in the discussions of CAC and BLSC, an important result of locating the new pool in the central area is that it has attracted users from existing pools. This could have been foreseen and might have been used as one criterion for assessing possible sites, especially since the choice of a central location has made some of the older pools less viable than before. It is therefore interesting to try to estimate this effect by examining the catchments of all the old pools under various assumptions before and after the imaginary siting of a new pool at the three possible locations under examination.

Under the assumption of Thiessen catchments, when the new pool is located at Holyrood Park Road, the catchment of Infirmary Street is reduced by 35.4%, Warrender by 40.6% and Portobello by 19.4% (Table 8.4 (b)). Since the pool at Infirmary Street already had the smallest Thiessen catchment area, its viability would seem to be more threatened by this choice of site because Warrender is still left (in theory) with a substantial tributary population (Table 8.4 (a)).

A pool at Meadowbank would have a similar effect on Infirmary Street, a slightly greater effect on Portobello, and would reduce the catchment of the pool in Leith by 30%. On the other hand the pool at Roseburn leaves all these catchments untouched but steals most of Dalry's patrons (-77.3%) and a third of Glenogle's. Despite this loss Glenogle still retains a moderately large catchment (Table 8.4 (a)) and, though Dalry then has one of the smaller catchments in Table 8.4, it does not have the very smallest.

When overlapping catchments are used, the new pool reduces the

TABLE 8.1

Effect of locating 'central pool' at various sites on older pools, assuming Thiessen catchments

(a) Population of catchments

	<u>Six old pools</u>	<u>Location of additional pool</u>		
		<u>1. Holyrood Park Road</u>	<u>2. Roseburn</u>	<u>3. Meadowbank</u>
Dalry	139,168	139,168	31,639	139,168
Glenogle	83,121	83,121	55,477	83,121
Infirmery Street	37,763	24,379	37,763	25,822
Leith	48,663	48,663	48,663	35,051
Portobello	43,816	35,314	43,816	32,868
Warrender	100,621	59,810	99,665	100,621
New pool		62,697	136,129	36,501

(b) Percentage change in population of catchments

	<u>Location of additional pool</u>		
	<u>1. Holyrood Park Road</u>	<u>2. Roseburn</u>	<u>3. Meadowbank</u>
Dalry	0	-77.3	0
Glenogle	0	-33.3	0
Infirmery Street	-35.4	0	-31.6
Leith	0	0	-30.0
Portobello	-19.4	0	-25.0
Warrender	-40.6	-1.0	0

catchments of all the older pools to some extent, no matter where it is located (Table 8.5), whereas with Thiessen catchments only three of the older pools were affected by any particular new location. When b_s is 0.50 the broad pattern of results is similar to the preceding one in many respects: a pool at HPR substantially reduces the catchments of the pools at Infirmary Street, Portobello and Warrender; a pool at Roseburn reduces the catchments of Dalry and Glenogle by more than a quarter of their size; a pool at Meadowbank draws most of its users from the former catchments of Portobello, Leith and Infirmary Street. One striking difference, however, is that the pool at Roseburn reduces the catchment of Warrender by over 19% whereas it had only a negligible effect on Warrender in the previous analysis (Table 8.4).

A more general difference is that with overlapping catchments all the reductions are smaller in percentage terms than with Thiessen catchments. The effect of the Roseburn pool on the baths at Glenogle is a particularly notable instance, since the reduction is now 37.7% as compared to 77.3% previously.

When b_s is 0.25 there is an even greater tendency for the new sites to affect all the old pools more equally. Nevertheless, the relative magnitudes of the reductions are very similar to those just discussed. For instance, Roseburn's main impact is on the pools at Dalry, Glenogle and Warrender, reducing those catchments by 23.4%, 20.2% and 18.2% respectively.

A similar method can be used to assess the impact of the alternative sites on the index of use within each catchment. With Thiessen catchments the overall pattern of reductions is so similar to that for size of catchments that it is not necessary to show them

TABLE 8.5

Percentage reduction in size of catchments of older
pools with Commonwealth Pool located at various
sites, assuming overlapping catchments

	1. <u>Holyrood Park Road</u>	2. <u>Roseburn</u>	3. <u>Meadowbank</u>
(a) <u>$b_s = 0.25$</u>			
Dalry	12.7	23.4	8.4
Glenogle	11.4	20.2	10.2
Infirmary Street	15.9	15.0	13.0
Leith	12.6	12.2	15.7
Portobello	17.7	5.8	19.9
Warrender	15.9	18.2	10.3
(b) <u>$b_s = 0.50$</u>			
Dalry	8.0	37.7	3.1
Glenogle	5.7	27.5	5.9
Infirmary Street	20.7	11.9	14.7
Leith	8.3	6.0	20.6
Portobello	18.1	0.8	23.8
Warrender	19.0	19.4	6.9

here but the pool at Dalry is not affected quite so badly by the pool sited at Roseburn. The reason is probably that it has a large concentration of population fairly close by in Gorgie - Dalry, Fountainbridge and Polwarth and therefore a relatively high population potential.

If overlapping catchments are used with $b = .25$ and $b_s = .50$, the results again suggest that the site at HPR particularly reduces the level of use at Infirmary Street (Table 8.6 (b)). It should be noted that Infirmary Street already had the lowest index of use in absolute terms, apart from Portobello. Moreover, although the index of use for the pool at Dalry is reduced by 28.8% with a new pool at Roseburn, in absolute terms, the pool at Dalry still enjoys then a higher level of use than Infirmary Street does even without the new pool sited at HPR. The same conclusion can be drawn when $b = .125$ and $b_s = .50$ (Table 8.6 (a)). Results are only shown for these particular values of b and b_s because these are the values which are likely to be closest to reality, representing as they do situations with moderately elastic demand but fairly compact catchments.

When attractiveness is taken into account, the disproportionate attraction of the new pool naturally reduces the levels of use at all the older pools to a much greater extent than in the preceding estimates (Table 8.7). In ordinal terms, however, the estimates are similar to the previous ones in that the greatest impact of the new pool at HPR is still on Infirmary Street, followed by Warrender, Portobello, Dalry, Leith and Glenogle in that order. The equivalent order for a pool at Roseburn is also the same as it was before the incorporation of attractiveness. In absolute terms the difference

TABLE 8.6

Percentage reduction in predicted level of use of
older pools with Commonwealth Pool located at
various sites, assuming overlapping catchments

	1. <u>Holyrood Park Road</u>	2. <u>Roseburn</u>	3. <u>Meadowbank</u>
(a)			
<u>$b = .125, b_s = .50$</u>			
Dalry	7.6	33.1	2.7
Glenogle	5.2	21.5	5.9
Infirmmary Street	21.6	7.8	15.5
Leith	7.2	3.2	20.9
Portobello	14.6	0.4	22.5
Warrender	19.4	14.8	6.3
(b)			
<u>$b = .25, b_s = .50$</u>			
Dalry	7.1	28.8	2.4
Glenogle	4.8	17.1	5.7
Infirmmary Street	22.2	5.5	15.1
Leith	6.4	2.0	20.3
Portobello	11.8	0.3	20.4
Warrender	19.2	11.6	5.5

TABLE 8.7

Percentage reduction in predicted level of use of
older pools with Commonwealth Pool located at
various sites, assuming overlapping catchments
and taking attraction into account

	1. <u>Holyrood Park Road</u>	2. <u>Roseburn</u>	3. <u>Meadowbank</u>
(a)			
<u>b = .125, b_s = .50</u>			
Dalry	26.7	62.7	10.4
Glenogle	19.8	46.9	19.8
Infirmary Street	51.0	21.6	38.4
Leith	25.5	10.6	50.4
Portobello	39.6	2.2	53.4
Warrender	47.9	36.3	20.5
Total old pools	34.5	33.8	29.6
(b)			
<u>b = .25, b_s = .50</u>			
Dalry	26.5	59.3	9.7
Glenogle	19.4	41.7	20.3
Infirmary Street	52.6	18.2	38.6
Leith	23.9	8.3	50.8
Portobello	34.4	1.6	50.2
Warrender	48.6	32.3	19.0
Total old pools	33.6	29.3	33.9

between the largest impacts and the smallest is now less, so the predicted effects are less unequal than in the previous estimates.

A further difference is that the reduction in predicted use at Dalry caused by a pool at Roseburn is now sufficiently great that the absolute level of use at Dalry is sometimes less than that at Infirmary Street with a new pool at HPR. This is true for both values of b_s when $b = .125$, but not when $b_s = .50$ and $b = .25$. When attractiveness is incorporated, a pool at Roseburn therefore tends to undermine the viability of that at Dalry to a greater extent.

When a general comparison of all the predicted impacts of the new pools is made from Table 8.7, the site at HPR has a clear disadvantage vis-à-vis that at Roseburn in that it affects all the old pools to a significant extent with the possible exception of Glenogle. On the other hand, the new pool at Roseburn has relatively little effect on the pools at Portobello and Leith and reduces the predicted use of the baths at Infirmary Street by only 19.8% or 19.4%, depending on the value of b .

It is also possible to assess the impact of a new pool by comparing the predicted level of use of all the old pools taken together. When $b = .125$ and $b_s = .50$ a new pool at HPR reduces the level of use of all the old pools by 34.5%, marginally more than the 33.8% drop caused by a pool at Roseburn. When $b = .125$ and $b_s = .25$ the respective falls are 44.5% and 42.4%. With $b = .25$ and $b_s = .50$ the corresponding figures are 33.6% and 29.3%. Though these values of b and b_s are arbitrary their range is sufficiently great to be reasonably sure of including the true value. Hence on the basis of these figures a pool at Roseburn not only stimulates more 'new' demand, but appears to

capture slightly less demand from the older pools. It is interesting to note in passing that when $b = .125$ and $b_s = .50$, a pool located at Meadowbank causes an even smaller drop than one at Roseburn.

In conclusion, this examination of the predicted impact of a new pool on the older ones does not furnish decisive evidence for strongly preferring any one of the three sites considered to the other two. If attraction is not considered it does suggest that one undesirable result of siting the new pool at HPR would be to erode the catchment area of the pool in Infirmary Street, which already has a smaller theoretical catchment and level of use than most of the older pools. Although a new pool at Roseburn would substantially reduce the sizes of catchments and levels of use of the pools at Dalry and to a lesser extent, Glenogle, both of these pools would still appear to retain much larger catchments and levels of use than Infirmary Street.

The assessments based on attractiveness also suggest that a pool at Roseburn would have affected fewer of the older pools and that it would draw marginally fewer users from the old pools. Its impact on the pool at Dalry, however, might be as great or greater than that of the pool at HPR on Infirmary Street and Warrender.

(b) Actual impact of the new pool

Information is available on admissions to the various pools for the period before and after the Commonwealth Pool was opened, so it is possible to assess its actual impact at the site in HPR. By comparing admissions at each pool in the year before the Commonwealth Pool was opened, 1968-69, with admissions in the second full year of its opening, 1971-72, the percentage reduction in use can be calculated (Table 8.8). Information on admissions is unfortunately

TABLE 8.8Admissions to pools for the years1968-69 and 1971-72

	<u>1968-69</u>	<u>1971-72</u>	<u>1971-72 as</u> <u>% of 1968-</u> <u>69</u>	<u>Percentage</u> <u>change</u>
Dalry	193,509	117,828	60.9	-39.10
Glenogle	176,825	128,510	72.6	-27.32
Infirmery Street	189,517	114,254	60.3	-39.70
Leith	188,467	137,787	73.2	-26.90
Portobello	348,860	222,035	63.6	-36.40
Warrender	221,008	128,698	58.2	-41.80
Commonwealth Pool		907,168		
Total	1,318,186	1,756,280	133.2	+ 33.20

not available after 1971-72 in a form which permits this type of comparison so readily, therefore a thorough comparison of actual and predicted reductions in use over a longer period cannot easily be carried out here.

The relative drop in admissions is greatest for Warrender and Infirmary Street, as expected, but also for Dalry, which was not expected from the preceding analysis. The drop in admissions at Dalry, however, is not entirely surprising since the Dalry pool is only 2.75 km. from the Commonwealth Pool. As expected, Portobello's admissions also fell very substantially.

The most striking feature of these results, however, is the way the Commonwealth Pool has affected all the other pools (Table 8.8). Even the relatively distant pools of Leith and Glenogle both suffer a drop of over 25%. The obvious reason is that the Commonwealth Pool is so much more attractive that it is able to draw users from a much wider catchment than the other pools. Using again the analogy of a cone, its catchment has a much broader base and more gently sloping sides than the other catchments. A further point of interest is the way total admissions to all pools rises after the construction of the Commonwealth Pool (Table 8.8), further confirmation that demand for swimming is spatially elastic and therefore responds to an improvement in access. Part of the increase, however, must also be stimulated by the sheer attractiveness of the new pool and cannot therefore be attributed to spatial elasticity.

The actual declines in use can be employed as a means of testing the models from which the predictions were made. By comparing the predicted reductions in Tables 8.6 and 8.7 with the

actual figures in Table 8.8, it is obvious that the model which included attractiveness is much more successful with respect to the actual magnitudes of change. Nevertheless, the model which did not include attractiveness is often broadly correct as to which pools are affected most and which least in relation to particular sites for the new pool.

The accuracy of the model with attractiveness in predicting the relative levels of change can be gauged by calculating the Pearson correlation co-efficient between the actual and predicted figures. When $b = .125$ and $b_s = .50$ the correlation is 0.78 which represents an explained variation of about 61%. The correlation for predictions based on $b = .125$ and $b_s = .25$ and on $b = .25$ and $b_s = .50$ are very similar. Although it has to be borne in mind that these correlations are only based on six pairs of values, the predictions are surprisingly accurate, especially since the values of b and b_s were predetermined and somewhat arbitrary. In fact by calibrating the model to the actual figures concerned a closer correlation could be obtained.

The most striking inaccuracy of the estimates given in Table 8.7 is that the predictions are too high for the pools most affected by the Commonwealth Pool (Infirmary Street and Warrender) and too low for the least affected, Glenogle. For instance, when $b = .125$ and $b_s = .50$ the predicted reductions at Infirmary Street and Warrender are 51.0% and 47.9% respectively as opposed to 39.7% and 41.8% in reality. On the other hand the predicted reduction at Glenogle is only 19.8% whereas in reality it was much higher, 27.3%. The predictions for Leith and Portobello are quite accurate but the discrepancy at Dalry is quite large: a real drop of some 39% compared to the estimate of roughly 26.5%. This surprisingly large

reduction at Dalry may be partly explained by the rather good bus connections which exist between the Commonwealth Pool, Gorgie-Dalry and areas to the west of Gorgie-Dalry.

On the whole these discrepancies suggest that the model exaggerates the impact of the Commonwealth Pool on the two pools nearest to it, but tends to underestimate the effect on pools further away, especially Glenogle and Dalry. In short the range of predicted values is greater than the range of actual values. Part of the explanation is probably that, as noted earlier, the real attractiveness of the Commonwealth Pool is higher than the value used in the model. Part of it may also be that the value of .5 for b_s gives a catchment which is too compact and that a slightly lower value would yield better predictions.

The predictions made for the site at Roseburn cannot, of course, be tested but the relative accuracy of those for HPR lends more confidence to the model on which both are based. Even if some allowance is made for the tendency of the model to underestimate the effect of the new pool on the more distant pools at Leith and Portobello, it nevertheless seems reasonably clear that a new pool at Roseburn would have affected fewer of the old pools so adversely as the pool at Holyrood Park Road has done.

In fact the viability of the pool in Infirmary Street has since been questioned. In 1977 the district council discussed a proposal to close it but, after lobbying by some of the pool's users, the proposal was eventually rejected. Thus, with the wisdom of hindsight it seems a pity that more attention was not paid by CAC and BLSC to the likely effect of the new pool on the older ones.

Conclusion

In summarising this examination of the decision to locate the Commonwealth Pool, it must first be emphasised that, as spatial interaction models, the models used or implicit in LOCHWISP are relatively unsophisticated and their use within an algorithm for locating central facilities is still somewhat experimental. Nevertheless they do allow accessibility to be treated more clearly and realistically than it was treated in the discussions of CAC and BLSC preceding the decision, a point substantiated by their limited success in prediction.

From the discussions of these two bodies it would appear that accessibility was not given sufficient weight as a crucial factor influencing both where the users of the new pool would come from and what the implications would be for the older pools. For all their limitations, the models used are able to throw some light on both points, especially when attractiveness is included. Had more weight been given to accessibility, CAC, BLSC and the city council might well have persisted more firmly with their earlier decision to locate the new pool at Roseburn. However, this point is only of historical interest now. More important for our present purpose is the fact that the models employed have yielded some worthwhile insights into the problem and provided some interesting criticisms of the decision made. The models therefore appear to be worth retaining for future refinement. It might also be suggested that when such a large investment in a central facility is being made, a local authority might well find it worthwhile to test possible sites by such a relatively inexpensive device as a location model.

Against the view that Roseburn would have been a better site it

can be argued that at its present location the Commonwealth Pool has undoubtedly been a great success. In fact it was claimed at one time that it was the only swimming pool in Britain with more than one million admissions annually. However, with the same design and facilities it would probably have been at least as successful at Roseburn; the evidence presented here suggests that it would have drawn even more use there and done so at less expense to some of the existing pools.

CHAPTER 9

The planning and location of district pools

Introduction

Discussion of plans for the provision of district pools in Edinburgh has a long and complex history. As early as 1958, BLSC and CAC agreed in principle to construct four district pools in suburban areas in addition to the new central pool. At the present time only one of such pools has actually been built. Yet in 1972 the general policy of building four district pools was still being affirmed.

It is difficult to summarise the complex way in which the priority given to various sites seems to have changed over this long period. In analysing these changes the main sources of information have again been the minutes of CAC and BLSC and, after the reorganisation of local government implemented in 1975, the minutes of the Recreation and Leisure Committee (RLC) of the City of Edinburgh District Council and one of its sub-committees, the Physical Recreation Sub-Committee (PRSC). Of course, such minutes are only a partial record of the real progress of making decisions. The picture presented here therefore is liable to be distorted by whatever bias the minutes have or by whatever is omitted from them. Despite their limitations, these minutes do seem to indicate clearly which sites were in serious contention at any time and they sometimes provide useful information about the difficulties of proceeding at specific sites. Their main drawback is that arguments comparing the relative merits of rival sites are either mentioned very briefly or omitted altogether.

From December 1958 to early 1972, the principle that four district pools should eventually be built was broadly reaffirmed on several occasions. During that period there was little disagreement about the four districts to receive pools, these being Clermiston, Crewe Toll, Comiston and Sighthill/Wester Hailes (Figure 8.1). However, there were disagreements about the exact siting of the pool within Comiston and the location of the pool for Sighthill was eventually moved about three quarters of a kilometre from the College of Commerce in Sighthill to a site adjoining the new community school for Wester Hailes, which in fact lies between Wester Hailes and Sighthill.

The main concern in the present chapter is firstly to explain why one of these locations got priority over the others and in particular to ask what influence the locational characteristics of the sites had on the final decisions. Secondly, we wish to ascertain whether the plans prepared and decisions made were efficient in spatial terms.

History of the main decisions regarding district pools

At various times, designs were prepared for pools at Comiston, Crewe Toll and Sighthill/Wester Hailes. Indeed, on several occasions the minutes give the impression that construction of one or more of these three is about to start in a matter of months. Yet CAC and BLSC never seem to have debated explicitly which should come first, tending instead to reaffirm that the first would be followed relatively quickly by at least two others. An example is the meeting of BLSC on 2nd March 1971 at which it was agreed that construction should begin on a pool at Comiston in 1971-72 to be followed by Sighthill (1973-74), Clermiston (1975-76) and Crewe

Toll (1977-78). This apparent confidence in a sustained programme of construction now seems very unrealistic and must have been questionable at the time, given the repeated delays from 1958 onwards. It may be noted here that, although the pool at Clermiston came third on this occasion, it never seems to have reached the design stage and, unlike the other three, never came first at any time.

In general the minutes of CAC and BLSC in the early 1960s give the impression that the pool at Comiston would be built first. As the preceding example shows, this was still the case as late as March 1971. However, early in 1971 the Education Committee (EC) of Edinburgh Corporation suggested that a public swimming pool should be one of the facilities included in the new community school at Wester Hailes. A joint sub-committee of EC and CAC, formed to examine this proposal, reported favourably. The proposal was finally approved by Edinburgh Corporation itself on 28th October 1971, despite the opposition of BLSC (and to some extent CAC) which still held that a pool at Comiston should come first. CAC and BLSC also argued that a site adjacent to Edinburgh College of Commerce or a site adjoining Broomhouse Primary School (Figure 8.1) would be better locations for a pool in the Sighthill/Wester Hailes area because both are more central within the pool's likely catchment. This argument was of no avail: work started on the site at Wester Hailes in 1975 and the pool eventually opened in 1978.

Having been forced to accept that the first district pool would be at Wester Hailes, BLSC then debated where the second should be located. On 30th May 1972 the General Manager of Baths and Laundries presented a report which included information on the 'service area, population statistics and distance from the nearest

existing pool' for sites at Comiston, Clermiston and Crewe Toll. The Sub-Committee then voted by eight votes to two in favour of Crewe Toll as opposed to Clermiston but the minutes give no indication of what influence the information presented had on the decision reached. Accepting this recommendation, Edinburgh Corporation designated Crewe Toll as the next site for a district pool on 22nd June 1972.

During the next 18 months arrangements were actually made and completed for the transfer of some four acres of land within the grounds of Telford College at Crewe Toll from the city's Education Account to the Civic Amenities Account to allow the pool to be built there. Thus, at that time there clearly was a genuine intention to build a pool at Crewe Toll. In fact on one occasion the City Architect reported that the pool at Crewe Toll was to be started in June 1974 and completed in November 1975 at a cost of £885,000. Yet the pool was never started and in the present financial climate, it is unlikely to be started for a considerable time.

One immediate reason for the delay was that it was agreed shortly afterwards to build a dry sports centre adjacent to the pool. Accordingly CAC asked for more land to be transferred from the Education Account but the Education Committee refused. CAC thought that the area already transferred was insufficient for both purposes and further delay was occasioned while this matter was investigated. Eventually in 1976, studies showed there was enough room for both facilities, but by then the financial climate had changed.

Explanation of priorities

To understand what factors influence the choice of locations for pools it is worth asking why the first district pool was built at Wester Hailes, as opposed to the other three locations, and why Crewe Toll was selected as second, though never built. It may also be worth asking why so little of the original plan for constructing four pools has been fulfilled during the twenty years after it was accepted in principle.

One important reason why Wester Hailes took precedence is that it offered the possibility of integrating the pool within a school complex. The advantages of such an arrangement lie in sharing a source of heat and in providing ready access for use during school hours or immediately after by one of the age groups with the greatest frequency of use. From the early 1960s the Scottish Education Department encouraged this type of integration and it is possible that this may have made it easier for Edinburgh Corporation to secure government approval of the financial arrangements, though this point was never mentioned in the Minutes of BLSC or CAC. The fact that Wester Hailes, a council estate built mainly in the late 1960s and very early 1970s, was known to be poorly provided with social amenities of all kinds, particularly for recreation, may also have had some influence with the joint sub-committee and with the Corporation.

The possibility of integrated provision helps to explain why Wester Hailes took priority over Comiston and Clermiston since no plans existed to build new secondary schools or colleges there. However, it may not explain why Wester Hailes had priority over Crewe Toll as the latter pool was to be built within the grounds of

Telford College of Further Education where some degree of integration would have been possible. Moreover, if the lack of amenities in Wester Hailes was an important argument, an advocate of the pool at Crewe Toll could probably have made an equally strong case in terms of the lack of such provision in the nearby districts of Pilton, Drylaw and Muirhouse. It seems therefore that the argument for preferring Wester Hailes to Crewe Toll on grounds of integration or under provision of amenities is not necessarily very persuasive.

It may be worth recalling that in March 1971 the pool at Sighthill came second in the programme of construction outlined by BLSC and CAC whereas the pool at Crewe Toll only came fourth. Though the minutes give no indication of why Sighthill came before Crewe Toll or why Comiston came first, the fact that it had earlier been given precedence may have been a marginal advantage.

Yet there is an obvious and very sound reason for giving priority to a pool at Sighthill or Wester Hailes, namely that both sites are over 4.5 km from the nearest pool at Dalry, whereas the site at Crewe Toll is only about 1.7 km from the pool at Glenogle. Although this point may conceivably have been important in the discussions of CAC, BLSC and the joint sub-committee, curiously there is no mention of it in the minutes. Thus, although it seems relatively clear why Wester Hailes had priority over Comiston and Clermiston, it is much less clear what actually persuaded the various bodies concerned to give it priority over Crewe Toll.

In fact, it is difficult to assess how strong this relative preference actually was. It is possible that in 1972-73 many members of CAC and BLSC were fairly confident that the pool at Crewe Toll would be started fairly soon and might even be open before the

pool at Wester Hailes. Therefore it may have seemed there was little need to assess which pool was more important. Consequently, the pool at Crewe Toll may have been a somewhat fortuitous casualty of the more stringent financial climate prevailing from 1973 onwards, simply because of the extra delay involved in starting it. This is probably part of the explanation why the pool at Crewe Toll was not built, but some of the relevant minutes in 1971 and 1972 would suggest that the site in Wester Hailes or Sighthill had a definite precedence.

It is also interesting to ask why Crewe Toll had priority over Clermiston, Comiston and other locations as the next site after Wester Hailes. As noted earlier the possibility of some degree of integration with Telford College was probably an important advantage which the sites at Clermiston and Comiston did not possess. It also seems likely that the report presented to BLSC in May 1971 on the 'service area, population statistics and distance from the nearest existing pool' for the three sites favoured Crewe Toll, because this meeting voted decisively for the latter site after discussing the report. If the report favoured Crewe Toll, it could not have been in terms of distance because Crewe Toll is only 1.7 km from the pool at Glenogle whereas Comiston and Clermiston are 3 km and 4.4 km respectively from their nearest pools, Warrender and Dalry.

Since it has not been possible to obtain a copy of this report so far, the exact basis on which the service areas and population statistics were calculated by the General Manager of Baths and Laundries is not clear. In trying to estimate the size of catchments of the three sites using LOCHWISP it was assumed that a

pool was already in existence at Wester Hailes. When Thiessen catchment areas were used, Crewe Toll was found to have a population of 68,437 in its tributary area whereas Clermiston had about 5,000 fewer and Comiston had only some 36,000. For both Thiessen and overlapping catchments, Crewe Toll also had more population within a radius of 2.5 km but only marginally more in the latter case. If it can be assumed that the evidence in the report was broadly similar to these results, it seems likely that the decision was influenced by this particular evidence. The site at Crewe Toll is also a particularly good one for bus routes, whereas Clermiston is less well connected. This point is not mentioned, however, in the minutes of BLSC or CAC.

Contrary to this evidence, when cost of travel and level of use in the whole system are computed under various assumptions, the results marginally favour Clermiston. This is largely because Crewe Toll draws more users from existing pools, particularly Glenogle, whereas Clermiston tends to stimulate more new demand. The priority given to Crewe Toll can therefore be criticised on these grounds.

Since the existing pools in 1971 were not well located to serve the city as a whole and since the four district pools would have greatly improved this situation, the most significant question in some respects may be why only one of the four was built rather than why it was located at Wester Hailes. Several reasons can be given. The cost of the Commonwealth Pool, £1.7 million, left much less room for other large recreation projects in the Corporation's budget during the period 1967-71. Moreover, the size and significance of this project and of the stadium at

Meadowbank, built for the Commonwealth Games in 1970, may have absorbed some of the energy and attention as well as the capital which might have gone into pursuing the plan for district pools. Conversely, after 1970 the success of the Commonwealth Pool probably encouraged the Corporation to build more pools. In fact one of the advocates of the pool at Wester Hailes mentioned this point as part of his case.

A further reason for delay may be that the emphasis on integrating the district pools in colleges or secondary schools necessitated co-operation between CAC and EC. Since these committees sometimes had different priorities regarding the various sites, the immediate result may have been to retard implementation of plans. Immediately after the reorganisation of local government in 1974-75 the fact that education was the responsibility of Lothian Regional Council, whereas recreation was shared between the latter and the City of Edinburgh District Council, may well have made co-operation more difficult. Since 1973-74, however, the main reason for delay has probably been the reduction in expenditure on public services by central government.

A factor which is much harder to assess is the strength of the political will behind the plan, particularly in the early 1960s when the Corporation did not have the financial burden of the Commonwealth Pool and the general financial climate seemed to be rosier. Some letters in the 'Scotsman' at the time give the impression that some ratepayers were opposed in general to large sums of money being spent on swimming pools. It is possible that the political composition of Edinburgh's Council, which almost invariably has a Conservative majority, tends to make the council less willing to embark on costly facilities for the community which

will raise the rates. It is, nevertheless, impossible to assess this factor properly without comparing patterns of expenditure on recreation across a wide range of local authorities of different political complexions.

To summarise the foregoing discussion, it is possible to outline a series of factors which help to explain both the general delay and the priority eventually given to Wester Hailes but it is difficult to assess how much weight should be assigned to each of them, especially in the case of the latter discussion. The cost of the Commonwealth Pool certainly delayed plans for district pools in the period roughly from 1967 to 1970. After 1973 cuts in public expenditure apparently made it difficult to proceed with new pools and this is probably one of the main reasons why the pool at Crewe Toll was not built.

It is more difficult, however, to say why the various bodies concerned eventually gave priority to a pool at Wester Hailes. The opportunity of integrating the pool within a new community school was certainly one important reason but it does not satisfactorily explain why the pool was built at Wester Hailes rather than Crewe Toll. There are sound reasons in terms of population and distance why the Sighthill area is probably a better location for a pool but, to judge by the various minutes consulted, this had little influence on the decision taken. Other factors may have played a role but they are hard to assess. It may even be the case that the local councillor for Wester Hailes was a very persuasive advocate. Whatever the reasons, it is fairly clear that before 1972 little attempt was made to assess the potential sites explicitly in terms of access to population. Some attempts were made to assess such

factors after that date but it is difficult to know how satisfactory they were. It is therefore particularly interesting to examine these sites more rigorously using the methods developed earlier.

Modification of population base

In the previous chapter the analysis of the decision to locate the Commonwealth Pool was undertaken using the population in 1971 of the area within the boundary of the local authority responsible for that decision, Edinburgh Corporation. Though the decision to construct a new pool at Wester Hailes was also taken by Edinburgh Corporation, in examining this decision it seemed more appropriate to use the boundary of the new local authority which assumed responsibility in 1975, the City of Edinburgh District Council (CEDC). It may be worth noting that the new authority actually started work in 1974, though it did not take over from the old until 1975. The new authority's boundary includes an extensive area to the west of the old, mostly used for agriculture but containing a few small towns and outlying suburbs of the city with a total population of 22,812 people.

The main reason for using the new boundary is that the new pool is located on the extreme western edge of the old authority's area and must therefore draw some of its users from the newly incorporated areas. Moreover, the boundary of the new area was known in 1973 and data on its population was available by late 1974, so the population of the new area could have been taken into account, at least in theory, by both old and new authorities before construction was started, late in 1975. Since the new pool was constructed as a joint project involving both the Recreation and Leisure Committee of CEDC and the Education Committee of Lothian Regional Council, it may be inferred that both new authorities either

approved the choice of location, or were not sufficiently opposed to cancel it.

When the census for 1971 was taken the new council estate at Wester Hailes was still under construction and contained only some 4,000 of the 15,000 - 18,000 people it is now estimated to house. Since Wester Hailes is undoubtedly the main area in the city where the data for 1971 is now inaccurate and since it also forms a very important part of the new pool's likely catchment, an attempt was made to obtain a more accurate estimate of the distribution of population in the area as it was in 1975. Information on the number of houses in each part of the estate had been obtained by John McCalden, a postgraduate student at Edinburgh, for use in his own research and this information was kindly made available. The sections of the estate were small enough to be assigned reasonably accurately to grid squares and the population of each grid square was then estimated by assuming that each house contained three persons, a rough figure recommended as appropriate by the Housing Department of the council. Allowance was made for the fact that some cells contained population which had already been counted in the census.

Since the new population in Wester Hailes largely came from older areas in the inner part of the city and since there was no means of estimating where exactly they had come from, the modified data base inevitably counts some people twice. For the present purpose this is a less serious error than omitting some 10,000 people from the area immediately next to the new pool because the error of double counting will be spread over a relatively large number of grid cells in the older part of the city. The catchments

of the older pools will therefore tend to be overestimated but this overestimate will be shared between the pools at Leith, Infirmary Street, Dalry, Holyrood Park Road and, to a lesser extent, Warrender.

Strategy of locational analysis

If LOCHWISP is used to obtain efficient locations for additional pools, the number of additional pools specified should obviously be fairly realistic. The question therefore arises of what that number should be. The long standing plan of BLSC and CAC was to construct four district pools; in the early 1970s two pools eventually received final approval; in the end only one of these was built.

It would be possible to test the original set of four locations to see whether a better set of four could be found on some or all criteria. Similarly, the two selected could be examined in the same manner or we could simply assess the choice of Wester Hailes as a location for one additional pool. Another possibility would be to use an incremental approach, initially finding the best additional location given the seven pools in existence in 1974, then finding the best location for a ninth pool with the eight built at Wester Hailes.

It should be noted that the two best locations obtained from an incremental approach might be different and somewhat less efficient than the best solution obtained by locating two pools simultaneously. Since the question of the best set of four pools is now largely an academic one, it has been ignored in the following analysis, which concentrates on the question of the best locations for one and two additional pools.

Analysis of Wester Hailes as the site for an additional pool

Computations, similar to those used to evaluate possible sites for the Commonwealth Pool, were undertaken using each of four original sites for district pools as the location of an eighth pool. Three other locations were also used in these trials: the site at Wester Hailes where it was actually built; the site next to Broomhouse School, recommended by the city's General Manager of Baths and Laundries; and a site at Gracemount on the southern edge of the city which was proposed in 1973 as the location for a sports centre including a 'leisure pool'. This last proposal was accepted in principle but never put into effect. To facilitate comparison between the actual site and its possible rivals, the values obtained for each site were expressed as percentages of the equivalent values for Wester Hailes. As in the previous chapter values for the whole system of catchments rather than for the new facility itself are discussed first.

When these comparisons are made on the basis of Thiessen catchments, the various results for the three sites within the general area of Sighthill are found to be consistently better than those for the other four locations in terms of travel cost, level of use and maximum distance travelled, often by a large margin (Table 9.1). In fact none of the values for Clermiston, Crewe Toll, Comiston or Gracemount is ever better than the equivalent value for any of the three locations in Sighthill. Clermiston is clearly the best of the former group of four locations, yielding fairly good values for both travel cost and level of use. Crewe Toll and Comiston are much less efficient in terms of travel cost and are both somewhat less effective in raising the overall

TABLE 9.1

Indices of accessibility with a new district pool at specific sites, relative to the values for W. Hailes, assuming Thiessen catchments

	<u>Seven pools</u>	<u>W. Hailes</u>	<u>Eighth pool constructed at</u>				<u>Crewe T.</u>	<u>Com.</u>	<u>Grace.</u>
			<u>Sight.</u>	<u>Broom.</u>	<u>Cler.</u>				
<u>Cost of travel (AT)</u>	128.7	100	97.5	97.7	104.8	118.6	114.4	120.1	
<u>Maximum distance travelled</u>	120.0	100	95.1	100.4	102.2	119.6	119.6	120.0	
<u>Use (TRIPS)</u>									
(1) b = .125	91.3	100	100.8	101.3	98.4	95.2	96.2	95.0	
(2) b = .25	88.1	100	101.1	102.3	98.0	94.5	95.4	94.4	
(3) b = .50	86.7	100	100.7	103.0	98.1	95.6	96.0	95.6	
<u>Sight.:</u> Sighthill (next to College of Commerce)						Crewe T.:	Crewe Toll		
<u>Broom.:</u> Broomhouse (next to Broomhouse Primary School)						Com.:	Comiston		
<u>Cler.:</u> or Clerm.:						Grace.:	Gracemount		

level of use in the system. The three locations in Sighthill give fairly similar values, the actual site being consistently the least efficient, by a small margin. It is interesting that the site at Broomhouse School is the best of the three, being slightly better than that at the College of Commerce on nearly all the measures shown, except maximum distance travelled.

As before, the measure of attractiveness used for the new facility was based on the surface area of the pool itself. The pool built at Wester Hailes is a 'leisure pool' with a clover shape instead of the traditional rectangular shape of conventional pools. Its surface area has been estimated as 1.9 times that of the older pools at Warrender and Dalry, but this probably underestimates its attractiveness since it provides such amenities as a toddlers' pool, a diving pool and a cafeteria which the older pools do not have. Also it is part of a recreation centre offering a wide range of activities. It should be noted that this value of attractiveness was applied to all the sites examined as alternatives to Wester Hailes.

When attractiveness is taken into account and overlapping catchments are used, the results are similar to the previous ones in that the three locations at Sighthill are always better than those for the other four sites (Table 9.2). The difference is again a very substantial one in most cases, though it is usually somewhat smaller than with Thiessen catchments. Clermiston is an exception to this general pattern because it appears to be at more of a disadvantage vis-à-vis Wester Hailes in terms of cost of travel and level of use than when Thiessen catchments are used.

There is much less difference between the three locations in Sighthill than when Thiessen catchments are used. As a result the

TABLE 9.2

Indices of accessibility with a new pool at specific sites relative to values for W. Hailes, using overlapping catchments and attractiveness

	<u>Seven</u> <u>pools</u>	<u>W. Hailes</u>	<u>Sight.</u>	<u>Broom.</u>	<u>Cler.</u>	<u>Crewe T.</u>	<u>Com.</u>	<u>Grace.</u>
<u>Cost of travel</u>								
(1) $b_s = .25$	116.7	100	102.4	103.2	105.3	112.7	112.1	115.9
(2) $b_s = .50$	126.0	100	99.2	99.5	105.5	118.3	114.8	121.4
<u>Use (TRIPS)</u>								
<u>$b = .125$</u>								
(1) $b_s = .25$	90.6	100	98.7	98.6	96.8	93.2	93.4	91.9
(2) $b_s = .50$	89.2	100	100.4	100.8	97.5	93.1	94.3	92.3
<u>$b = .25$</u>								
(1) $b_s = .25$	85.8	100	97.9	98.2	95.1	90.1	90.2	88.6
(2) $b_s = .50$	84.7	100	100.0	101.8	96.6	91.0	92.4	90.5

site at Wester Hailes is placed at less of a disadvantage compared to the other two sites on all measures (Table 9.2). Indeed, when $b_s = .25$, the cost of travel to the site at Wester Hailes is lowest of the three probably because this location then draws proportionally more of the population in outlying areas to the west and therefore has more effect on aggregate travel. Though there is little to choose between these three sites on all other measures, with the more realistic value for $b_s = 0.5$, the site at Broomhouse appears to stimulate marginally more demand in the system as a whole.

Results were also computed for overlapping catchments without taking attractiveness into account. These results place the site at Wester Hailes at a marginally greater disadvantage compared to the two nearby locations but place the other four sites at less disadvantage. In this instance, the omission of attractiveness therefore reduces the difference between a set of relatively good locations and a set of poorer ones but also increases the differences observed within a group of good locations which are fairly near each other. The explanation seems to be that when a facility placed at a good location in the centre of an unserved area is given a high value for attractiveness, it draws more nearby users than when its attraction is equal to the other facilities. The local component of its service area then has a greater resemblance to a Thiessen catchment and this tends to reduce travel cost and increase use. Even if the facility is displaced a short distance, its attraction still ensures that this favourable effect is maintained. Conversely, if the same facility is placed at a much poorer location its attractiveness draws more users from relatively long distances and this has an adverse effect on both objectives. As a very broad

rule, the introduction of attractiveness draws proportionately more users to a good location but also draws more to an inefficient location and so increases the difference between the values computed for the two.

Of the four locations outside Sighthill, Clermiston is easily the best and Gracemount consistently the worst; Comiston tends to be marginally better than Crewe Toll (Table 9.2). Crewe Toll was originally thought to be a rather good location because of the large population in the adjoining estates of Muirhouse, Pilton and Drylaw, but this site makes less contribution to increasing use and reducing travel cost than expected, mainly because it is only 1.7 km from the existing pool at Glenogle.

Equivalent values were also computed for NBU with both types of catchment. Since the conclusions drawn from these values were very similar to those for the indices of use, the values of NBU are not shown here. It should be noted, however, that the numerical differences between the sites in terms of NBU were greater than for level of use.

If the index values specifically associated with each site are compared without taking attraction into account, the difference between the various locations at Sighthill appears much greater than in the preceding analysis (Table 9.3). Thus the facility at Broomhouse School enjoys a catchment area between 15% and 20% larger than that at Wester Hailes and between 8% and 10% larger than that at the College of Commerce in Sighthill, depending on the assumption used. Furthermore, the level of use of the pool at Broomhouse is between 18% and 24% better than that of the site in Wester Hailes - an appreciable difference.

TABLE 9.3

Facility index values for various sites, given as percentage of values at W. Hailes (without considering attractiveness)

	<u>W. Hailes</u>	<u>Sight.</u>	<u>Broom.</u>	<u>Cler.</u>	<u>Crewe T.</u>	<u>Com.</u>	<u>Grace.</u>
(a) <u>Thiessen catchments</u>							
Size of catchment	100	103.5	115.6	106.1	85.4	82.0	32.3
(b) <u>Overlapping catchments</u>							
<u>Size of catchment</u>							
(1) $b_s = .25$	100	108.4	116.7	101.3	92.1	90.4	52.0
(2) $b_s = .50$	100	110.6	120.7	103.1	83.6	86.8	40.3
<u>Level of Use</u>							
(1) $b = .125; b_s = .25$	100	108.4	118.3	93.5	82.4	77.4	50.3
(2) $b = .125; b_s = .50$	100	111.2	123.0	95.3	79.5	77.8	48.6
(3) $b = .25; b_s = .25$	100	106.6	117.5	88.2	75.2	68.6	52.6
(4) $b = .25; b_s = .50$	100	110.2	123.9	90.7	77.5	72.4	55.4

When attraction is taken into account the results for individual facilities are more complex (Table 9.4). When $b_s = .50$ Broomhouse has a significantly bigger catchment than Wester Hailes, and a substantially higher level of use. However, when $b_s = .25$, the reverse is true and Wester Hailes is clearly the best of the three locations at Sighthill. Since $b_s = .50$ is probably a more realistic value, these results for overlapping catchments could be taken as some vindication of the judgement of the city's General Manager of Baths and Laundries, who suggested that Broomhouse was a better location with respect to the catchment formed by the western part of the city.

The preceding discussion has been concerned with the objectives computed for the whole system and for the new pool itself. An analysis was also carried out of the effect each of the sites has on the levels of use of the seven older pools (including the pool at HPR), using attractiveness and allowing catchments to overlap. The results of this analysis are shown for certain values of b and b_s and for selected sites in Table 9.5. With the standard range of values for b and b_s , with few exceptions the old pools always maintain at least 70% of their former levels of use, no matter which of the seven sites is considered.

The most striking exceptions occur when a new pool at Crewe Toll reduces the level of use at Glenogle to 56.8% and 60.3% of its former volume with $b_s = .50$ for $b = .125$ and $.25$ respectively. The other exceptions occur when new pools at Sighthill College and at Broomhouse reduce the predicted use of Dalry to 69.3% and 64.3% respectively of its original value with $b = .125$ and $b_s = .50$. It should also be noted that a pool at Clermiston would have a roughly

TABLE 9.4

Facility index values for various sites, given as percentage of values at W. Hailes with attractiveness considered

	<u>W. Hailes</u>	<u>Sight.</u>	<u>Broom.</u>	<u>Clerm.</u>	<u>Crewe T.</u>	<u>Com.</u>	<u>Grace.</u>
(b) <u>Overlapping catchments</u>							
<u>Size of catchment</u>							
(1) $b_s = .25$	100	78.7	83.0	72.9	61.3	57.3	28.6
(2) $b_s = .50$	100	96.8	104.8	89.0	69.9	69.1	26.9
<u>Level of Use (TRIPS)</u>							
(1) $b = .125; b_s = .25$	100	81.7	87.3	70.5	57.6	50.2	28.2
(2) $b = .125; b_s = .50$	100	100.1	109.8	85.1	69.0	64.0	34.6
(3) $b = .25; b_s = .25$	100	82.7	89.3	69.1	54.5	45.6	30.5
(4) $b = .25; b_s = .50$	100	101.7	113.5	83.4	69.2	61.4	41.7

Predicted level of use of older pools with new pool at various sites as a percentage of the original value, using overlapping catchments and attractiveness

(a) $b = .125; b_s = .25$

	<u>W. Hailes</u>	<u>Sight.</u>	<u>Broom.</u>	<u>Clerm.</u>	<u>Crewe T.</u>	<u>Com.</u>	<u>Grace.</u>
Dalry	81.8	79.6	76.6	80.7	81.8	82.8	95.1
Glenogle	90.4	83.3	86.3	82.4	74.4	91.5	97.2
Inf. St.	92.6	91.5	89.8	91.1	87.2	90.6	94.5
Leith	97.0	96.3	95.3	93.5	84.1	96.3	97.6
Portobello	99.2	99.0	98.6	98.9	96.4	97.3	92.3
HPR	93.6	93.0	91.4	93.6	91.0	90.1	91.0
Warrender	86.5	85.3	82.3	88.1	87.9	83.1	92.4

(b) $b = .125; b_s = .50$

Dalry	72.9	69.3	64.3	72.1	78.6	76.1	98.1
Glenogle	91.6	89.0	86.9	76.3	56.8	95.0	99.6
Inf. St.	93.6	92.7	91.1	92.3	88.8	92.0	96.5
Leith	99.3	99.0	98.8	97.3	86.8	99.4	99.8
Portobello	99.9	100.0	100.0	100.0	99.6	99.7	96.7
HPR	95.3	94.9	93.6	96.0	95.2	90.9	88.0
Warrender	82.3	81.0	77.0	87.0	90.8	73.7	93.4

Predicted level of use of older pools with new pool at various sites as a percentage of the original value, using overlapping catchments and attractiveness

(c) $b = .25$; $b_s = .25$

	<u>W. Hailes</u>	<u>Sight.</u>	<u>Broom.</u>	<u>Clerm.</u>	<u>Crewe T.</u>	<u>Com.</u>	<u>Grace.</u>
Dalry	87.4	85.4	81.9	85.8	84.0	85.5	95.9
Glenogle	94.2	92.5	90.6	86.7	75.0	93.8	97.8
Inf. St.	96.0	95.3	93.8	94.7	89.5	93.1	95.6
Leith	98.6	98.1	94.4	96.1	85.9	97.5	98.1
Portobello	99.6	99.5	99.2	99.4	97.3	98.0	93.3
HPR	96.7	96.3	95.0	96.5	93.2	92.6	92.1
Warrender	91.5	90.6	87.9	92.4	90.4	85.8	93.5

(d) $b = .25$; $b_s = .50$

Dalry	83.5	80.6	75.5	82.2	83.4	82.7	98.7
Glenogle	96.1	94.4	92.7	83.9	60.3	97.3	99.8
Inf. St.	97.5	97.0	95.9	96.6	92.2	95.6	98.0
Leith	99.8	99.7	99.6	98.8	89.5	99.7	99.8
Portobello	100.0	100.0	100.0	100.0	99.7	99.8	97.7
HPR	98.2	98.0	97.2	98.5	97.1	94.5	90.9
Warrender	90.7	89.9	86.4	93.5	94.0	80.7	95.2

equal impact on the pools at Glenogle and Dalry; a pool at Comiston would reduce the level of use at both Dalry and Warrender to roughly three quarters of their previous levels. In almost all other instances, however, the older pools retain 80% of their previous use.

An assessment of the impact of a particular site on all seven pools can be obtained by comparing their joint level of use before and after the imaginary location of the new pools. The site with the most adverse effect is again Crewe Toll. For instance, when $b = .125$ and $b_s = .50$ a new pool at Crewe Toll reduces the aggregate use of the old pools to 85.1% of its former level whereas the corresponding figures for Wester Hailes, Sighthill College and Broomhouse are 90.3%, 89.0% and 86.8% (Table 9.6). On this point the site at Wester Hailes has a clear advantage over Broomhouse and in fact enjoys some advantage over all the other sites. However, since the reduction in use of all the older pools never exceeds 15%, this effect is less important than in the case of the Commonwealth Pool, where the predicted reductions were often over 30% and the actual decrease was 33.2%.

In summary, it can be said that none of the sites seems to pose a major threat to the level of use at any of the seven existing pools, with the exception of the site at Crewe Toll. Of the various locations in Sighthill/Wester Hailes, the site at Wester Hailes has the advantage that it reduces the estimated use of all the older pools and Dalry in particular to a lesser extent than does the site at Broomhouse.

The various sites considered for a new district pool have now been compared in terms of their effect both on the system as a whole

TABLE 9.6

Predicted impact of new district pool on
seven existing pools using overlapping
catchments and attractiveness

- (a) Level of use of all old pools after location of a new pool at
various sites as a percentage of original level

<u>Location of new pool</u>	<u>b = .125</u>	<u>b = .125</u>	<u>b = .25</u>	<u>b = .25</u>
	<u>b_s = .25</u>	<u>b_s = .50</u>	<u>b_s = .25</u>	<u>b_s = .50</u>
W. Hailes	91.5	90.3	95.0	95.1
Sighthill C.	90.4	89.0	94.2	94.3
Broomhouse S.	88.6	86.8	92.5	92.5
Clermiston	90.1	88.3	93.5	93.3
Crewe Toll	86.8	85.1	88.7	88.0
Comiston	89.6	88.9	92.1	93.0
Gracemount	93.4	94.6	94.5	96.3

- (b) Increase in overall use stimulated by new pool as a percentage
of predicted level of use at new pool itself

<u>Location of new pool</u>	<u>b = .125</u>	<u>b = .125</u>	<u>b = .25</u>	<u>b = .25</u>
	<u>b_s = .25</u>	<u>b_s = .50</u>	<u>b_s = .25</u>	<u>b_s = .50</u>
W. Hailes	54.9	55.6	76.9	78.9
Sighthill C.	53.4	54.5	75.0	77.4
Broomhouse S.	49.8	51.3	71.0	73.9
Clermiston	46.3	46.2	67.5	69.1
Crewe Toll	23.3	25.1	37.8	41.1
Comiston	28.2	36.8	46.0	59.1
Gracemount	18.9	41.2	42.8	67.0

and on the existing pools and in terms of their own role in the system. From this comparison, it must be concluded that, if one pool was to be built, the best location was one of the three sites in Sighthill/Wester Hailes. When attractiveness is not considered, the best of these three locations in terms of travel cost and level of use in the whole system is at Broomhouse School. When attractiveness is included in the analysis, there is very little to choose between the three. The site at Wester Hailes has the advantage of a smaller impact on the old pools than Broomhouse but has the disadvantage of producing a marginally smaller level of use in the system than Broomhouse does when $b_s = .50$.

Thus, in terms of locational criteria, the process of decision making yielded a more efficient result in this instance than in the case of the Commonwealth Pool. Since locational criteria hardly seem to have had any explicit influence on the final choice, the efficiency of the outcome may be purely a matter of chance. It is also possible, however, that intuitive but sound assessments of the basic factors of population and distance made as the decision makers studied a map of the city somehow guided the final choice. This is not impossible because relatively good solutions to certain location/allocation problems can be obtained by studying a detailed map of the distribution of population. This is often the case with problems involving Thiessen catchments but intuitive judgements are much less reliable when attractiveness and overlapping catchments are involved, especially if the impact of a new facility on existing ones has to be estimated.

Nevertheless, the major factor leading to the choice of the site at Wester Hailes was probably the opportunity to integrate the

new pool in a secondary school complex. If this factor is given priority over locational criteria, then some price may be paid by users in terms of higher costs of travel and lower utilisation. The preceding analysis suggests that if the new pool had been built at Crewe Toll, this price would have been quite high, but that it was not very significant for the actual location chosen at Wester Hailes.

Plans relating to the proposed pool at Comiston

As late as 2nd March 1971 BLSC agreed that the first of the new district pools to be constructed should be at Comiston followed by Sighthill, Clermiston and Crewe Toll in that order. Since the preceding results make Comiston a much less efficient location than both Sighthill and Clermiston, the priority given to Comiston then and in the preceding years seems misplaced. The opening of a pool at Comiston, however, was part of a wider plan. Reaffirming earlier decisions, BLSC and CAC also agreed in early 1971 that when the pool at Comiston opened, the pools at Dalry, Infirmary Street and Warrender should close. It is therefore instructive to try to estimate what effect a pool at Comiston might have had on the older pools.

If Thiessen catchments are used, Dalry and Warrender are left with 74.8% and 68.5% respectively of their original levels of use when $b = .125$ (Table 9.7 (a)). When $b = .25$ both retain an even higher proportion of the original figure. When overlapping catchments are used and the effect of attractiveness included, the levels of use at Dalry and Warrender always exceed 80% of their original values, except when $b = .125$ and $b_s = .50$ (Table 9.7 (b)). In the latter instance they only drop to 76.1% and 73.7% respectively.

TABLE 9.7

Predicted level of use at seven existing
pools after location of a pool at Comiston
as a percentage of the original value

(a) Thiessen catchments

	<u>b = .125</u>	<u>b = .25</u>
Dalry	74.8	85.7
Glenogle	100	100
Inf. St.	100	100
Leith	100	100
Portobello	100	100
HPR	97.4	98.7
Warrender	68.5	79.3

(b) Using attractiveness and overlapping catchments

	<u>b = .125</u>	<u>b = .125</u>	<u>b = .25</u>	<u>b = .25</u>
	<u>b_s = .25</u>	<u>b_s = .50</u>	<u>b_s = .25</u>	<u>b_s = .50</u>
Dalry	82.8	76.1	85.5	82.7
Glenogle	91.5	95.0	93.8	97.3
Inf. St.	90.6	92.0	93.1	95.6
Leith	96.3	99.4	97.5	99.7
Portobello	97.3	99.7	98.0	99.8
HPR	90.1	90.9	92.6	94.5
Warrender	83.1	73.7	85.8	80.7

As the pool at Infirmary Street always keeps at least 90% of its original level of use, it is particularly difficult to see why the construction of a pool at Comiston should be linked to the closure of this pool.

Since the model used to make these predictions was moderately successful in estimating the impact of the Commonwealth Pool on the older pools, it seems reasonable to conclude that a pool at Comiston would have had little effect on the levels of use at Dalry and Warrender and almost no effect on admissions at Infirmary Street. Hence it is rather difficult to understand why this plan was approved, but it may have been assumed implicitly that the new pool would mainly capture users from existing pools. There may therefore have been a failure to fully appreciate the point that, when demand is moderately elastic, a well located pool will stimulate a substantial volume of new demand. Even with such a mildly elastic demand as that given by $b = .125$, nearly 37% of the predicted admissions at Comiston in fact represent new demand (Table 9.6 (b)).

As noted earlier, there may have been a case for closing the pool at Infirmary Street, but it could not be based on the pool proposed for Comiston. Furthermore, there can be little doubt that all of the closures would have been opposed fiercely and, quite possibly, successfully by their users.

The fact that such an unsatisfactory plan was approved and reaffirmed suggests that locational models and algorithms such as LOCHWISP could be helpful to planners and decision makers in testing plans and providing rough guidelines as to efficient locations. At a later date, in fact, Lothian Regional Council

commissioned a study of swimming pools which made use of location/allocation models (Cargill and Hodgart, 1978). The latter study was based on the NORLOC algorithm, however, and the limitations of this algorithm have been mentioned already, the most notable being its assumption of inelastic demand and its inability to treat overlapping catchments.

Choice of two locations for district pools

It is possible to test the selection eventually made in 1972-73 of two locations for district pools, namely W. Hailes and Crewe Toll. One way to examine this choice is by comparing it with other pairs of possible locations for the eighth and ninth pools in the city (Table 9.8). In this analysis it was assumed that both the eighth and ninth pools would have the same attractiveness as the pool at W. Hailes, a value of 1.9. Of the various pairs of locations tested, the combinations which came out particularly well were:

- (a) Broomhouse and Davidsons Mains;
- (b) Broomhouse and Muirhouse;
- (c) W. Hailes and Muirhouse.

This was true both in trials using attractiveness and without it. In terms of the total cost of travel and level of use these pairs tended to be between 1.5% and 6% better than W. Hailes and Crewe Toll when attractiveness is used, depending on the values of b and b_s (Table 9.8). Both pairs involving Broomhouse have similar costs of travel, but Broomhouse and Muirhouse yield slightly higher levels of use, partly because the location in Muirhouse is fairly central to the large concentration of

TABLE 9.8

Indices of accessibility based on overlapping catchments and attractiveness for two new pools at selected sites given as percentages of the values with new pools at W. Hailes and Crewe Toll

		<u>W. Hailes</u> & <u>Crewe T.</u>	<u>W. Hailes</u> & <u>Muirh.</u>	<u>Broom.</u> & <u>D.Mains</u>	<u>Broom.</u> & <u>Muirh.</u>	<u>Slate.</u> & <u>Clerm.</u>	<u>Colin.</u> & <u>Clerm.</u>	<u>Colin.</u> & <u>Corst.</u>
<u>Cost of travel</u>								
(a)	$b_s = .25$	100	97.9	97.9	98.2	99.9	99.8	100.0
(b)	$b_s = .50$	100	96.1	95.1	95.0	98.3	98.1	99.1
<u>Level of use</u>								
(1)	$b = .125; b_s = .25$	100	101.5	101.4	101.8	99.6	99.4	99.2
	$b_s = .50$	100	102.1	102.5	103.1	100.3	100.1	99.6
(2)	$b = .25; b_s = .25$	100	103.3	102.3	104.0	98.7	98.2	97.6
	$b_s = .50$	100	104.2	104.1	106.1	100.0	99.2	98.5
Muirh.:	Muirhouse		Colin.:	Colinton				
D. Mains:	Davidsons Mains		Corst.:	Corstorphine				
Slate.:	Slateford							

population in the neighbouring estates of Drylaw and Pilton and in Muirhouse itself. The pair consisting of W. Hailes and Muirhouse is also consistently superior to W. Hailes and Crewe Toll and is nearly as efficient as Broomhouse and Muirhouse. It is therefore clear that the disadvantages of W. Hailes and Crewe Toll as a pair are mainly attributable to the location of Crewe Toll.

The pairs selected were also compared in terms of both their impact on existing pools and of their success in stimulating new demand as opposed to drawing demand away from the existing pools (Table 9.9). On the former criterion the difference between the pairs is not great but W. Hailes and Crewe Toll erode the demand of the older pools to a slightly greater extent than all the other pairs selected. Naturally the pairs furthest from the centre of city (Colinton and Clermiston; Colinton and Corstorphine) have least effect on the older pools. As far as stimulating new demand is concerned, all the other pairs are noticeably more successful than W. Hailes and Crewe Toll. When $b = .125$ and $b_s = .50$ the pairs involving Broomhouse respectively yield 6.5% and 5.3% more and the pair consisting of W. Hailes and Muirhouse 6.6% more new demand than W. Hailes and Crewe Toll. The fact that W. Hailes and Crewe Toll are at a disadvantage on all of these comparisons is again largely due to Crewe Toll being relatively close to the existing pool at Glenogle.

Thus the three pairs of locations singled out earlier are better on all aspects of the preceding analysis than the pair of locations eventually chosen by the city council. As can be seen from the tables, however, the difference is usually not great.

Predicted impact of two new district pools on seven existing pools using overlapping catchments and attractiveness

<u>W. Hailes & Crewe T.</u>	<u>W. Hailes & Muirh.</u>	<u>Broom. & D.Mains</u>	<u>Broom. & Muirh.</u>	<u>Slate. & Clerm.</u>	<u>Colin. & Clerm.</u>	<u>Colin. & Corst.</u>
---------------------------------	-------------------------------	-----------------------------	----------------------------	----------------------------	----------------------------	----------------------------

(a) Level of use of seven existing pools

$b = .125; b_s = .25$	80.9	82.9	82.1	80.9	81.7	83.9	84.3
$b_s = .50$	78.4	80.9	80.5	78.7	80.2	82.6	83.7
$b = .25; b_s = .25$	85.1	87.7	87.1	85.8	86.5	88.8	89.2
$b_s = .50$	84.5	87.7	87.6	85.8	87.2	89.6	90.5

(b) Increase in use stimulated by new pools as a percentage of predicted level of use at new pools themselves

$b = .125; b_s = .25$	41.1	47.0	45.5	44.7	41.3	44.1	44.1
$b_s = .50$	43.4	50.0	49.9	48.7	46.1	48.9	49.8
$b = .25; b_s = .25$	59.8	68.1	66.0	65.5	60.5	64.1	64.0
$b_s = .50$	62.7	71.6	71.5	70.4	67.0	70.7	71.6

It can therefore be argued that W. Hailes and Crewe Toll are quite satisfactory as a pair of locations. Nevertheless, a location more central to the main concentration of population in the north west of the city and further from Glenogle would clearly be better than Crewe Toll as the northern partner of the pair.

Conclusion

One of the main concerns of this thesis has been to develop methods for solving problems involving spatially elastic demand and overlapping catchment areas. It is therefore worth asking whether the results obtained from these methods are very different from those derived from the simpler model based on inelastic demand and Thiessen catchments. When attractiveness was not allowed to vary and overlapping catchments were used, regardless of the particular objective the rank order of possible sites in terms of their efficiency for one additional pool was virtually the same as that for Thiessen catchments, given in Table 9.1. For instance, Broomhouse was always the best site for all values of b and b_g . In the specific context of choosing a location for one extra facility of the same attractiveness as all the existing ones, it could therefore be argued that the more complex method was not really justified.

Yet, even in this context, the simpler model provides rather inaccurate estimates of the impact of a new facility on existing ones, partly because it assumes that no new demand is created by the extra facility. Moreover, as Table 8.4 (b) indicates, it invariably predicts that only the catchments of the three immediately neighbouring facilities will be affected whereas the more complex model makes the more realistic prediction that all

the older facilities will be affected to some extent (Table 8.6).

When attractiveness was used, the correspondence between the results from the simpler and more complex models was not so close. Thus, whereas Broomhouse always yields a higher level of overall use than all the other sites in Table 9.1 irrespective of the value of b , in Table 9.2 the level of demand associated with a new pool at Broomhouse is highest when $b_s = .50$ but not when $b_s = .25$. Similarly, of all the possible new locations in Table 9.3 Broomhouse always enjoys the highest individual level of use whether Thiessen catchments are used or not, but when attractiveness is used (Table 9.4) its predicted admissions are much lower than those for W. Hailes with $b_s = .25$.

Thus the value of the more complex model mainly lies in estimating the impact of a new facility more accurately. Nevertheless, a further reason for using the more complex model is that, when attractiveness is used, assessments of possible sites can be somewhat different from those based on Thiessen catchments and inelastic demand.

The preceding examination of plans for district pools has shown that the location finally used at W. Hailes is quite an efficient one whether viewed as the site for a single additional pool or as one of a pair of sites to be added to the existing seven. Not all the plans which gained the approval of CAC and BLSC, however, were so sound. The long standing plan to make Comiston the first new district pool and to close the pools at Dalry, Warrender and Infirmary Street when it opened was clearly ill considered.

The limitations of the models used have to be borne in mind

in considering the results discussed in this chapter and the previous one. To obtain more accurate estimates and predictions from the models several improvements could be made.

First, accurate values of the parameters, b and b_s are needed and this requires a carefully designed survey. It would also be desirable to find out from this survey whether different values of b and b_s should be applied to different types of pool. Second, an investigation of the relation between attractiveness and level of use is also needed. From this research it might be possible to devise more appropriate measures of attractiveness for use in the spatial interaction model. Third, the model could be given a more flexible structure which would allow a wider range of relevant information to be taken into account, for instance bus routes, age structure and level of car ownership.

It may also be worth noting that the model used does not consider trips made to a pool from place of work or made as part of the journey between home and school or home and shopping centre. Such journeys are probably more important at Infirmary Street, lying near the centre of the city, than elsewhere, but they may well form a certain proportion of the journeys made to the pools at Leith, Dalry and also to the Commonwealth Pool. Preliminary work carried out by Currie (1977) seems to imply that such journeys do not form a very important part of the total admissions at any of the pools. Nevertheless, some investigation of the extent to which pools are used by nearby schools and what factors influence the level of this use could provide a further basis for

refining the model. This point of criticism can be seen as part of the wider problem that models of facility location treat one service in isolation from other services. For swimming pools this may be justified to some extent, but work is needed to determine which services should be treated together and which can be viewed as relatively independent of each other.

Since this research has been concerned with decisions made by a local authority, the boundaries of that authority provided a suitable, though not ideal, areal framework. However, the boundary of the study area does affect the results obtained, particularly when demand is assumed to be inelastic. Suppose, for the sake of argument, that the population of an area some distance to the west of Edinburgh happened to fall within the district's administrative boundary and that, for this population, the swimming pool at Broxburn in W. Lothian District was much closer than the pool at W. Hailes. With an inelastic model this area would then be assumed to contribute the same amount of demand to pools in Edinburgh as population in the central areas of the city. With a model based on elastic demand, the demand contributed would be attenuated by the distance involved but again no account would be taken of the fact that most of the area's population would prefer to use the pool at Broxburn.

In estimating the impact of a new pool it would therefore be more realistic to ignore administrative boundaries and to draw the areal framework widely enough to include all pools which could be in competition with those immediately under examination. All areas whose population might use the latter pools to more than a

negligible extent should also be included. Since Edinburgh is a fairly compact urban unit separated from adjoining settlements, especially on the west, by its green belt, adherence to administrative boundaries is a less serious fault than it might be elsewhere. Nevertheless, if administrative boundaries were to be ignored, the pool at Broxburn and the pools at Bonnyrigg and Newtongrange in Midlothian should probably be included.

Without these refinements the results discussed in this chapter and the previous one are best regarded as first approximations derived from an experimental model. With a more sophisticated model the assessments made of various sites might be somewhat different. Nevertheless, the results derived from LOCHWISP helped to reveal very clearly the illogical nature of the plan to close the three older pools on the opening of the new pool at Comiston. Even if some allowance is made for more sophisticated models producing somewhat different results, it is unlikely that this particular conclusion will have to be revised.

SECTION V

Conclusion

CHAPTER 10

Conclusion

This thesis has been concerned with some of the concepts and models which are fundamental to the question of how public services which are supplied through a system of central facilities should be organised spatially. Essentially, the thesis has concentrated on two aspects of this question. First, what is the optimal number of facilities and, as a corollary, under what conditions should the service be supplied in a large number of small facilities as opposed to a few large ones? Second, how can locations be found for any given number of facilities which are both efficient and equitable? Whereas the contribution made to answering the former question is purely conceptual, the main thrust of the work on location has been to develop new methods, make them operational and apply them to an appropriate service.

Fundamental to both aspects has been the question of how users may benefit from improved access to facilities in terms of reduced costs of travel and a higher frequency of use. In fact, the degree to which demand is spatially elastic and the form of the demand cone is relevant to optimizing both the number of facilities and their location and therefore forms a link between these two sections of the work.

In discussing the optimal number of facilities, it was argued that for services where facilities tend to come in fixed sizes, including swimming, it was reasonable to assume that the cost of supplying the service was a linear function of the number of

centres, m . Since the various benefits to users of increasing the supply could also be directly related to m , it was possible to devise a framework in which the balance of costs and benefits could be assessed to determine optimal values for m . As a general conclusion, in these circumstances the more spatially elastic demand is, the lower the optimal value of m .

Conversely, however, in a particular situation where the existing distribution of facilities leaves some areas poorly supplied, the more spatially elastic is the demand, the more likely are new centres to be viable through stimulating previously untapped demand. If demand is quite elastic, in these circumstances a new facility could conceivably attract as much use as any of the existing ones while scarcely impairing their level of use. In both the general case and the latter one, the value of m which is taken as optimal will also depend on whether a 'marginal' or 'break-even' approach is used.

By using a similar framework, it was also shown that a particularly interesting situation occurs where a service is not tied to fixed sizes of facility but is free to organise a given capacity in several smaller or a few larger facilities without increasing the total costs of supply (i.e. the case with spatially flexible supply). In such circumstances the solution with the highest feasible value of m is then the most efficient and most equitable arrangement within the terms of the model.

These models made use of such traditional assumptions as the isotropic plane; also, they lack a time dimension and their assumptions about supply costs are rather simplistic. Because of

these restrictive assumptions their conclusions must be regarded as rather naive. Despite their simplicity, the models did point to a few significant gaps in the existing state of knowledge about the spatial interplay of supply and demand. The most important of these gaps are the lack of information on which services have spatially elastic demand and on which enjoy a spatially flexible supply. The models also helped to provide a more explicitly spatial answer to the question posed by Teitz (1968) about how public services are able to inhibit or stimulate demand through their form of spatial organisation.

By their nature these models are difficult to investigate empirically. However, in directing attention to a number of questions on which empirical work is needed, it is possible they may eventually have some practical as well as conceptual value. In retrospect, these models and the argument supporting them in Chapter 3 would seem to be the most original part of the present work.

The main contribution made to the locational aspect of the problem has been to develop, test and apply new methods of solving a number of forms of the problem. Since the algorithms available at the time of writing could only solve problems where demand was inelastic and facilities had Thiessen service areas, the most valuable part of the locational work probably lies in developing a means of solving problems where demand is elastic and tributary areas overlap.

In technical terms the most difficult part of this work lay in building a spatial interaction model into the LOCHWISP algorithm so that catchments overlapped but the algorithm could

still search for better locations. It was shown in fact that a search conducted in these circumstances must progressively converge on better, though not necessarily optimal, locations. This discovery was quite unforeseen and has implications for future work.

Though this demonstration only concerned locations on a plane, it is possible to extend the argument involved to similar problems in a network. In this case the location stage of the algorithm finds nodes with lower travel cost or a higher level of use within each catchment. In the allocation stage each demand point then has more of its population assigned to centres which are now closer within the network and less to centres now further away. Thus, the allocation stage, like the location stage, can only lead to improvement in both goals; convergence must therefore occur. It may be worth noting here that convergence also seems to occur on a plane when attractiveness is included in the model, but this property has not yet been thoroughly tested.

For applied work it is essential that these locational models should be able to treat overlapping catchments and a range of objectives, including those derived from elastic demand. It is debateable, however, whether it is really necessary for practical purposes in planning to have algorithms which actually search. In assessing plans for the location of swimming pools in Chapters 8 and 9, the ability of LOCHWISP to explore space was rarely used; mostly, specific locations were evaluated without search. It was also noted that good solutions can often be selected visually by inspecting a detailed map of the distribution of population. This

makes the ability to search less valuable because it means that most of the relatively good locations for additional facilities can probably be selected visually and then evaluated simply as discrete points.

A computer program is still essential for this work because though good locations can be chosen by eye, as noted earlier, visual judgement is not likely to be accurate in predicting how much new demand will be stimulated or how new facilities will affect existing ones especially when the spatial interaction model used is a relatively complex one.

In defence of the effort to develop searching properties, it can be said that, since there seems to be little existing literature on the problems of maximising use and net benefit to users, they are of some technical and experimental interest in relation to these goals: they may also be of conceptual and heuristic interest in the teaching of locational analysis. Moreover, even in an applied context it may still be of some interest to know how far a good location departs from the optimal and in what direction the latter lies.

In comparing the two major aspects of the work, it can be said that the degree of spatial elasticity in demand has an important influence on the number of facilities which is optimal but less influence in determining the location which is optimal, especially for additional facility problems. The latter point is supported by the frequency with which the same locations were found in Chapter 7 as solutions to problems involving different objectives. Furthermore, in Chapters 8 and 9, locations which were found to be very efficient in terms of one goal were invariably found to be efficient

in terms of other objectives, particularly when attractiveness was not used.

In reviewing the empirical work, it should first be said that the main weakness was probably the failure to disaggregate the population of each grid cell by age, social composition, car ownership or by other variables likely to influence the use made of swimming pools. The model based on inelastic demand is, in fact, quite easy to disaggregate. Suppose that the frequency of use varies according to age but that, for all age groups, it is independent of travel cost. If we know the number of people in each age group in a particular cell and each group's frequency of use, we can calculate the number of journeys originating there simply by multiplying the numbers in each group by their respective frequencies. To obtain travel cost, these journeys can then be multiplied by the appropriate distance, as was done with total population in LOCHWISP. When searching, the total number of journeys to a particular centre from cell i , o_i say, then replaces p_i in iterative equations 6.6 and 6.7.

When demand is spatially elastic and each group has a different degree of elasticity, disaggregation is not so easy. If a realistic estimate is to be made of the demand emanating from i , for each group both the frequency of use at zero distance and the rate at which this frequency falls with distance must be known. Since b -values have not yet been measured for populations as a whole, let alone sub-groups within them, the difficulties of application are self-evident. Nevertheless, if the necessary information were available, it would be easy to simply evaluate a set of fixed locations for centres on a disaggregated basis as long as search

is not required.

It is interesting to consider whether it would be possible to search on a disaggregated basis when demand is elastic. If we wish to maximize use, the gradient of T_i becomes the sum of the gradient values for each sub-group at i , i.e. it consists of a series of terms, each of which is very similar to the expression derived for dT_i/dX_j in Appendix 2. From these gradients, expressions can be derived for X_j and Y_j which are equivalent to 6.10 and 6.11 but more complex because of the differing b -values for each sub-group. Although computational experience has not yet been obtained, it seems reasonable to expect that convergence will occur, as it did with 6.10 and 6.11. If so, it would be fairly easy to develop a disaggregated version of LOCHWISP able to maximize use.

It may be relevant to consider what effect disaggregation might have on the results obtained in Chapters 8 and 9. Previous studies suggest that, although all age groups participate to some extent, the group aged 10 to 14 visits pools between three and five times more frequently than the mean for all age groups (Cargill and Hodgart, 1978). The other age group which departs considerably from the mean is the group over 45 whose frequency of use seems to be roughly $\frac{1}{2}$ to $\frac{1}{3}$ of the mean. In terms of occupation the spectrum of users is also quite broad but the unskilled and semi-skilled generally appear to be under-represented.

If all grid cells had the same age and occupational profile, the results from a disaggregated model would of course be identical to those computed already. Thus, variables produced by disaggregation will alter the results only if sub-groups of the population are

spatially segregated to some degree. Since there are proportionally more young people in the new housing estates which ring the edge of the city and fewer old people in the inner part of the city, disaggregation by age will tend to move some of the optimal locations outwards. Since the estate at W. Hailes had a particularly high proportion of children in 1971, this location would then appear even more advantageous than before; conversely the site at Holyrood Park Road, being near the centre, would probably appear in a slightly less favourable light. However, any displacement of the optimal locations outwards is unlikely to be large, partly because age groups are generally not segregated very strongly.

Although social class groups are more segregated in Edinburgh than age groups, as noted earlier swimming has a relatively broad appeal to all occupational groups. It can therefore be argued that disaggregation by social class would have less effect than by age group. In fact, during the study of swimming pools prepared for Lothian Regional Council the effect of disaggregation was explored to a limited extent in the inelastic model by giving different weights to different age groups and to households with and without cars. In these experiments the best locations found for additional pools were almost always quite close to those found without disaggregation. Thus disaggregation is unlikely to alter substantially the results computed in Chapters 8 and 9. One of the main reasons is simply that the proportion of people in the relevant sub-groups varies less than the density of population in spatial terms and therefore exerts much less influence as a geographical variable. Nevertheless, for services with a narrower

spectrum of users than swimming, disaggregation could be more important.

The preceding discussion makes the implicit value judgement that groups should be weighted according to their actual level of use. It can be argued, however, that such a policy will merely reinforce existing inequalities by locating facilities nearer the groups which already enjoy a higher level of use by virtue of car ownership, higher income or more time for leisure. If the use of a service by a particular group is very much reduced through the cost or time involved in travel or even through particular distances being perceived as barriers, it can be argued on the basis of equity that facilities should be organised to encourage greater use by such groups. The latter policy may favour plans with a greater number of smaller facilities and will obviously tend to locate centres nearer disadvantaged groups, if these groups are spatially segregated. By neglecting disaggregation in the present work, we have implicitly given equal weights to all sub-groups of the population; this yields more equitable solutions than weighting by existing level of use. There is a general argument, however, for giving greater weight to disadvantaged groups, i.e. positive discrimination. Applied to swimming, this would tend to give more weight to areas with high proportions of unskilled occupations.

In fact, more case studies are needed to find out whether access and mobility are significant as factors explaining why some groups have higher levels of use or other groups have lower levels. In designing these studies, it would be desirable to allow for different groups having differing degrees of elasticity. Such

studies could help to mediate the conflict between a policy of weighting according to existing levels of use and a policy favouring the disadvantaged. Yet, since value judgements are ultimately involved, it may not be possible to resolve this debate by empirical means.

Despite the shortcomings of the empirical work, an interesting point to emerge was the importance of being able to estimate what effect a new facility has on the level of use at the existing ones, a point which seems to be largely neglected in existing work on facility location. This problem is likely to be quite common in reality and it is one where conventional models based on inelastic demand will be rather inaccurate in the case of a service with elastic demand. The method employed for this purpose achieved moderate success but there is considerable scope for improving the rather simple spatial interaction model on which it was based.

Till now, the application of models for locating facilities has been restricted by their inflexibility and by the difficulty of adapting them to the circumstances of particular services. By constructing more flexible models, it is hoped that this thesis has helped to make such models more suitable for application. Moreover, through concentrating on the spatial dimension of the problem, a framework has been developed which may allow information on relevant social and economic factors to be incorporated more effectively in future. It is hoped that this will make it easier to integrate better the social and spatial dimensions of the problem, a vital objective for future work.

From the discussion at the end of Chapter 9, it is clear that much work remains to be done in developing better locational models.

To judge by some of the decisions made about swimming pools in Edinburgh, for all their faults these models still represent a significant improvement on the way accessibility has been treated by public bodies in the past. It can only be hoped that the information which better models will yield may contribute to a clearer, more informed and more democratic debate about such decisions in the future.

REFERENCES

- Abernathy, W.J. and Hershey, J.C. 1971: A spatial allocation model for regional health service planning. Operations Research 19, 629-42.
- Abler, R., Adams, J.S. and Gould, P. 1971: Spatial organisation. New Jersey: Prentice Hall. (587 pp.)
- Austin, C.M., Smith, T.E. and Wolpert, J. 1970: The implementation of controversial facility-complex programs. Geographical Analysis 2(4), 315-29.
- Berry, B.J.L. 1967: Geography of market centers and retail distribution. New Jersey: Prentice Hall. (146 pp.)
- Berry, B.J.L., and Horton, F.E. 1970: Geographic perspectives on urban systems. N. Jersey: Prentice Hall. (564 pp.)
- Boots, B.N. 1974: Delaunay triangles: an alternative approach to point pattern analysis. Proceedings of the A.A.G. 6, 26-29.
- Cargill, S. and Hodgart, R.L. 1977: Computer atlas of Lothian Region; demographic, social, economic profiles. Dept. of Geography, Univ. of Edinburgh. Prepared for Lothian Regional Council, Dept. of Recreation and Leisure. (100 pp.)
- Cargill, S. and Hodgart, R.L. 1978: A strategy for the provision of swimming pools in the Lothian Region. Dept. of Geography, Univ. of Edinburgh. Research report prepared for Lothian Regional Council (30 pp.) (Reports were also prepared in the same series on bowling greens, golf courses and squash courts.)
- Christaller, W. 1966: Central places in S. Germany, (translated by C.W. Baskin). N. Jersey: Prentice Hall. (230 pp.)
- Chung, A.C. 1963: Linear programming. Columbus, Ohio: Merrill Books. (338 pp.)
- Church, R. and ReVelle, C. 1974: The maximal covering location problem. Papers of the Regional Science Association 32, 101-118.
- City of Edinburgh District Council Committee Minutes 1974-76: Recreation and Leisure Committee. Edinburgh.
- Clarke, T.J. 1975: Non-housing provision in new public sector housing schemes with particular reference to Scottish practice. University of Edinburgh, Dept. of Urban Design and Regional Planning, M.Phil. thesis.

- Collins, M.F. 1979: Large or local - some geographical issues in the planning and management of indoor sports centres and swimming pools in Britain. In Simhuber, K.A. and Jülg, F. (eds.), Studies in the geography of tourism and recreation (Wiener Geographische Schriften 53/54). Vienna: Verlag Ferdinand Hirt.
- Cooper, L. 1963: Location - allocation problems. Operations Research 11, 331-43.
- Cooper, L. 1964: Heuristic methods for location - allocation problems. Siam Review 6, 37-53.
- Cooper, L. 1967: Solutions of generalised locational equilibrium models. Journal of Regional Science 7, 1-18.
- Cooper, L. 1968: An extension of the generalised Weber problem. Journal of Regional Science 8, 181-98.
- Cooper, L. 1972: The transportation location problem. Operations Research 20, 94-108.
- Coppock, J.T. and Barritt, M.M. 1978: Application of computer techniques to information systems for planning. Edinburgh: Central Research Unit, Scottish Development Dept. (137 pp.)
- Coppock, J.T. and Duffield, B.S. 1975: Recreation in the Countryside. London: MacMillan. (262 pp.)
- Cox, K.R. 1965: The application of linear programming to geographic problems. Tijdschrift voor Economische en Sociale Geografie 56, 228-236.
- Currie, I.M. 1977: Edinburgh's indoor swimming baths: location, use and catchment areas. University of Edinburgh, Department of Geography, M.A. thesis. (70 pp.)
- Dear, M. 1974: A political theory of public facility location. Paper presented at the 21st North American Meetings of the Regional Science Association, Chicago.
- Dear, M. 1975: Locational strategies for controversial public facilities. Paper presented at the annual meeting of the Canadian Association of Geographers, Vancouver, May 1975.
- Edinburgh Corporation Committee Minutes, 1958-74: Civic Amenities Committee. Edinburgh.
- Eilon, S., Watson-Gandy, C.D.T. and Christofides, N. 1971: Distribution management: mathematical modelling and practical analysis. London: Griffin. (240 pp.)

Eilon, S. 1972: Goals and constraints in decision making. Operational Research Quarterly 23 (1), 3-15.

Gainey, G.J. 1977: Geographical constraints on the receipt of primary medical care: the spatial background to urban general practice. University of Edinburgh, Dept. of Geography, M.A. thesis.

Galbraith, J.K. 1974: Economics and the public purpose. London: Deutsch. (334 pp.)

Garrison, W.L. 1959: Spatial structure of the economy, II. Annals of the Association of American Geographers, 49(4), 471-82.

Goodchild, M.F. 1972: The trade area of a displaced hexagonal lattice point. Geographical Analysis 4(1), 105-7.

Gould, P.R. and Leinbach, T.R. 1966: An approach to the geographic assignment of hospital services. Tijdschrift Voor Economische en Sociale Geografie 57, 203-206.

Hakimi, S. 1964: Optimum locations of switching centers and the absolute centers and medians of a graph. Operations Research 12, 450-459.

Hakimi, S. 1965: Optimum distribution of switching centers in a communications network and some related graph-theoretic problems. Operations Research 13, 462-75.

Harris, C.D. 1954: The market as a factor in the localization of industry in the United States. Annals of the Association of American Geographers, 44, 315-48.

Harvey, D. 1973: Social justice and the city. London: Arnold. (336 pp.)

Hirsch, W.Z. 1968: The supply of urban public services. In Perloff, H. and Wingo L. (eds.). Issues in urban economics. Baltimore: The Johns Hopkins University Press. 477-525.

Hirsch, W.Z. 1973: Urban economics analysis. New York: McGraw Hill.

HMSO, 1969 and 1977: Scottish educational statistics. London: HMSO.

Hodgart, R.L. 1978: Optimizing access to public services. Progress in Human Geography 2, 17-48.

Hotelling, H. 1929: Stability in competition. Economic Journal 41, 41-57.

- Isard, W. 1960: Methods of regional analysis. Cambridge, Mass.: M.I.T. Press. (784 pp.)
- Jones, R. 1967: Central place theory and the hierarchy and location of shopping centres in a city: Edinburgh. In I.B.G. Study Group in Urban Geography. Aspects of central place theory and the city in developing countries.
- King, L.J. 1976: Alternatives to a positive economic geography. Annals of A.A.G. 66, 293-308.
- Kuhn, H.W. and Kuenne, R.E. 1962: An efficient algorithm for the numerical solution of the generalized Weber problem in spatial economics. Journal of Regional Science 4(2), 21-34.
- Lea, A.C. 1973: Location - allocation systems: an annotated bibliography. University of Toronto, Department of Geography, Discussion Paper 13. (273 pp.)
- Lee, T. 1957: On the relations between the school journey and social and emotional adjustment in rural infant children. The British Journal of Educational Psychology 27, 101-114.
- Levy, F., Meltsner, A.J. and Wildavsky, A. 1974: Urban outcomes; schools, streets and libraries. Berkeley: Univ. of California Press. (278 pp.)
- Levy, J. 1967: An extended theorem for location on a network. Operations Research Quarterly 18, 535-42.
- Lloyd, P.E. and Dicken, P. 1972: Location and space: a theoretical approach to economic geography. New York: Harper and Row.
- Loomba, N.P. 1964: Linear programming. New York: McGraw-Hill. (284 pp.)
- Lösch, A. 1954: The economics of location. New Haven: Yale University Press. (520 pp.)
- McAllister, D.M. 1976: Equity and efficiency in public facility location. Geographical Analysis 8(1), 47-63.
- McCalden, J. 1981: The spatial behaviour of dental patients in Edinburgh. University of Edinburgh, Dept. of Geography: Ph.D. thesis (in preparation).
- McGrew, J.C. and Monroe, C.B. 1975: Efficiency, equity and multiple facility location. Proceedings of the Association of American Geographers 7, 142-46.

Maranzana, F.E. 1964: On the location of supply points to minimise transport costs. Operational Research Quarterly 15, 261-270.

Marshall, J. U. 1969: The location of service towns. Toronto: Univ. of Toronto Press. (184 pp.)

Massam, B.H. 1972: The spatial structure of administrative systems. Commission on College Geography; Resource Paper 12. Association of American Geographers, Washington, DC.

Massam, B. 1975: Location and space in social administration. London: Arnold. (192 pp.)

Massam, B.H. 1980: Spatial search. Oxford: Pergamon. (295 pp.)

Miehle, W. 1958: Link length minimization in networks. Operations Research 6, 232-43.

Mishan, E.J. 1967: The costs of economic growth. Harmondsworth: Penguin. (240 pp.)

Morrill, R.L. 1974: Efficiency and equity of optimum location models. Antipode 6(1), 41-6.

Morrill, R.L. and Symons, J. 1977: Efficiency and equity aspects of optimum location. Geographical Analysis 9, 215-225.

Neft, D.S. 1966: Statistical analysis for areal distributions. Regional Science Research Institute, Monograph Series 2.

Nordbeck, S. and Rystedt, B. 1972: Computer cartography. Lund: Student-litteratur (315 pp.)

"
Oberg, S. 1976: Methods of describing physical access to supply points. Lund Studies in Geography, Series B, Human Geography 43. Lund: Gleerup. (141 pp.)

Olsson, G. 1974: Servitude and inequality in spatial planning: ideology and methodology in conflict. Antipode 6(1), 16-21.

Ostresh, L.M. 1973: An investigation of the multiple location - allocation problem. University of Iowa, Dept. of Geography: Ph.D. thesis. (213 pp.)

Rawls, J. 1972: A theory of justice. Oxford: Oxford University Press. (607 pp.)

Redcliffe-Maud Report (Royal Commission on Local Government in England 1966-69), 1969. Cmnd 4040. London: HMSO.

Redcliffe-Maud Report (Royal Commission on Local Government in England), 1968: Research studies 3: Economies of scale in local government services. London: HMSO. (184 pp.)

Redcliffe-Maud Report (Royal Commission on Local Government in England), 1968: Research studies 4: Performance and size of local education authorities. London: HMSO.

ReVelle, C.S. and Swain, R.W. 1970: Central facilities location. Geographical Analysis 2(1), 30-42.

ReVelle, C., Marks, D. and Liebman, J.C. 1970: An analysis of private and public sector location models. Management Science 16 (ii), 692-707.

ReVelle, C. and Church, R. 1976: A spatial model for the location construct of Teitz. Paper presented at Toronto meeting of Regional Science Association. (10 pp.)

Robertson, I.M.L. 1974: Road networks and the location of facilities. Environment and Planning A 6, 199-206.

Robertson, I.M.L. 1976: Accessibility to services in the Argyll District of Strathclyde: a locational model. Regional Studies 10(1), 89-95.

Rojeski, P. and ReVelle, C. 1970: Central facilities location under an investment constraint. Geographical Analysis 2(4), 343-60.

Rushton, G., Goodchild, M.F. and Ostresh, L.M. 1973: Computer programs for location/allocation problems. Iowa: Dept. of Geography, University of Iowa. (321 pp.)

Schlesinger, J.R. 1966: Organizational structures and planning. Rand Corporation.

Scott, A.J. 1969: On the optimal partitioning of spatially distributed point sets. In Scott, A.J., editor, Studies in Regional Science, London: Pion.

Scott, A.J. 1970: Location - allocation systems: a review. Geographical Analysis 2(2), 95-119.

Scott, A.J. 1971: Combinatorial planning, spatial analysis and planning. London: Methuen. (204 pp.)

Scott, A.J. 1975: Discrete dynamic locational systems: notes on a structural framework. In Karlqvist, A., Lundqvist, L. and Snickars, F. (eds.) Dynamic allocation of urban space. Farnborough, England: Saxon House. (383 pp.)

Scottish Sports Council, 1979: A study of sports centres and swimming pools. Edinburgh: Scottish Sports Council. (2 vols. 290 pp.)

Smolensky, E., Burton, R. and Tideman, N. 1970: The efficient provision of a local non-private good. Geographical Analysis 2(4), 330-42.

Simon, H.A. 1957: Models of man. New York: J. Wiley. (287 pp.)

Statistisk Sentralbyrå^o (Central Bureau of Statistics), 1969 and 1975: Statistisk Årbok (Statistical Yearbook). Oslo: Statistisk Sentralbyrå.

Taylor, P.J. 1971: Distance transformation and distance decay functions. Geographical Analysis 3(3), 221-38.

Taylor, P.J. 1975: Distance decay in spatial interactions. I.B.G. Study Group in quantitative methods, CATMOG 2. Norwich: Geo. Abstracts.

Taylor, P.J. 1977: Quantitative methods in geography. Boston: Houghton Mifflin. (386 pp.)

Teitz, M.B. 1968: Toward a theory of urban public facility location. Papers of the Regional Science Association 21, 35-51.

Teitz, M.B. and Bart, P. 1968: Heuristic methods for estimating the generalized vertex median of a weighted graph. Operations Research 16, 901-1092.

Tiebout, C.M. 1956: A pure theory of local expenditures. Journal of Political Economy 64, 416-24.

Toregas, C. and ReVelle, C. 1972: Optimal location under time or distance constraints. Papers of the Regional Science Association 28, 133-43.

Toregas, C. and ReVelle, C. 1973: Binary logic solutions to a class of location problems. Geographical Analysis 5, 145-55.

Törnqvist, G. 1963: Studier i industrilokalisering. Stockholm: Geografiska Institutionen vid Stockholms Universitet.

Törnqvist, G., Nordbeck, S., Rystedt, B. and Gould, P. 1971: Multiple location analysis. Lund Studies in Geography, Series C. General and Mathematical Geography 12, Lund: Gleerup.

Wagner, J.L. and Falkson, L.M. 1975: The optimal nodal location of public facilities with price sensitive demand. Geographical Analysis 7(1), 69-83.

Watson, J.W. 1976: Land use and Adam Smith. Scottish Geographical Magazine 92, 129-33.

Webber, M. 1973: Planning in an environment of change. In Cullingworth, J.B., ed., Problems of an urban society, vol. 3. London: Allen and Unwin. (195pp.)

Weiss, J.E. and Greenlick, M.R. 1970: Determinants of medical care utilization: the effect of social class and distance on contacts with the medical care system. Medical Care 8(6), 456-62.

Weiszfeld, E. 1937: Sur le point pour lequel la somme des distances de n points donnés est minimum. Tohuko Math Journal 43, 355-86.

Wilson, A.G. and Kirkby, M.J. 1975: Mathematics for geographers and planners. Oxford: Oxford University Press. (325 pp.)

Wolpert, J., Mumphrey, A. and Seley, J. 1972: Metropolitan neighborhoods: participation and conflict over change. Commission on College Geography, Resource Paper 16, Association of American Geographers, Washington, DC.

Wolpert, J., Dear, M. and Crawford, R. 1975: Satellite mental health facilities. Annals of Association of American Geographers 65 (1), 24-35.

Yeates, M. 1963: Hinterland delimitation: a distance minimising approach. Professional Geographer 15(6), 7-10.

APPENDIX 1Main computer programs used in the work(a) Programs written by others

<u>Program</u>	<u>Author</u>	<u>Function</u>
NORLOC	S. Nordbeck and B. Rystedt, Univ. of Lund, Sweden	Locates m centres on a plane to minimize cost of travel.
LAP	M. Goodchild, Univ. of W. Ontario, Canada	Locates m centres on a plane. Can take account of capacities and barriers and can al- locate by a spatial interaction model.
CAMGRID	J. Hotson, Univ. of Edinburgh	Draws choropleth maps by line printer.
GIMMS	T. Waugh, Univ. of Edinburgh	A general mapping system using the graph plotter for output.

(b) Programs written by the author

<u>Program</u>	<u>Function</u>
CONVERTS	Aggregates point data (e.g. census enumeration districts) to cells of variable size in a square grid using a form of output suitable for NORLOC.
PUNMAP ₄	Aggregates point data to cells in a square grid using a form of output suitable for input to LAP, CAMGRID and LOCHWISP.
GENEVA ₃	A short program which evaluates the level of demand on an isotropic plane for values of m from 1 up to a specified limit according to the value of b and degree of overlap selected. Can use equation 3.9 or 3.14 in this evaluation.
NETPL ₂₇	Computes level of demand, NBU MT and AT on an isotropic plane for values of m up to a specified limit according to the value of b . It also scales the results, if desired, and plots them on a graph.

<u>Program</u>	<u>Function</u>
DEMCONE	A short program which computes what proportion of the theoretical demand at one centre comes from within a specified radius, given the value of b or b_s . Based on equation 3.13.
LOCWISP	Locates m centres on a plane. Its main options and facilities are described in Chapter 7.

APPENDIX 2

Location of one centre on a plane to maximise use.

Let the number of trips made to a centre (X_j, Y_j) by the population, p_i , of any point i with co-ordinates (x_i, y_i) be T_i . Let T_i depend on the distance, d_i , of i from j so that

$$T_i = p_i e^{-bd_i}.$$

The way in which T_i changes as the position of the centre moves is given by the partial derivatives of T_i with respect to X_j and Y_j . These derivatives can be found by a method similar to that outlined in Chapter 6 for aggregate travel. Basically T_i is regarded as a function of the variables X_j and Y_j . We therefore wish to find $\frac{dT_i}{dX_j}$ and $\frac{dT_i}{dY_j}$.

$$\begin{aligned} \text{Let } u &= d_i^2 = (x_i - X_j)^2 + (y_i - Y_j)^2 \\ &= x_i^2 - 2x_i X_j + X_j^2 + y_i^2 - 2y_i Y_j + Y_j^2. \end{aligned}$$

$$\text{Then } T_i = p_i e^{-bu^{\frac{1}{2}}}$$

T_i is a function of u and u is a function of X_j . Hence by the chain rule

$$\frac{dT_i}{dX_j} = \frac{dT_i}{du} \cdot \frac{du}{dX_j}.$$

To find $\frac{dT_i}{du}$ we can use the general relation

$$\frac{dy}{dx} = y \frac{d(\ln y)}{dx}.$$

$$\begin{aligned} \text{Hence } \frac{dT_i}{du} &= p_i e^{-bu^{\frac{1}{2}}} \frac{d(-bu^{\frac{1}{2}})}{du} \\ &= p_i e^{-bd_i} (-b)^{\frac{1}{2}} d_i^{-1} \quad (\text{since } u^{-\frac{1}{2}} = d_i^{-1}) \end{aligned}$$

$$\text{Now } \frac{du}{dX_j} = -2x_i + 2X_j = (-2)(x_i - X_j)$$

$$\text{Therefore } \frac{dT_i}{du} \cdot \frac{du}{dX_j} = -\frac{1}{2} b p_i e^{-bd_i} d_i^{-1} (-2)(x_i - X_j)$$

$$\text{i.e. } \frac{dT_i}{dX_j} = b p_i (x_i - X_j) e^{-bd_i} d_i^{-1} .$$

If we consider i as one of n points surrounding the facility at j and Z_t denotes the total number of trips from these points to j then

$$\frac{dZ_t}{dX_j} = b \sum_{i=1}^n p_i (x_i - X_j) / e^{bd_i} d_i .$$

At the point where the total number of trips is at a maximum, this gradient must be 0 .

$$\text{i.e. } b \sum_{i=1}^n p_i (x_i - X_j) / e^{bd_i} d_i = 0$$

$$\therefore b \sum_{i=1}^n p_i x_i / e^{bd_i} d_i = b X_j \sum_{i=1}^n p_i / e^{bd_i} d_i$$

$$\therefore X_j = \frac{\sum_{i=1}^n \frac{p_i x_i}{e^{bd_i} d_i}}{\sum_{i=1}^n \frac{p_i}{e^{bd_i} d_i}} .$$

By symmetry

$$Y_j = \frac{\sum_{i=1}^n p_i y_i e^{-bd_i} d_i^{-1}}{\sum_{i=1}^n p_i e^{-bd_i} d_i^{-1}} .$$

This completes the derivation of equations 6.10 and 6.11 . Both have to be solved iteratively since d_i is unknown.

APPENDIX 3

Location of one centre to maximise NBU on a plane.

Using the same notation as in Chapter 3 and in the preceding appendix, let the net benefit to users at i of a facility at j be

$$\begin{aligned} \text{NBU}_i &= vT_i - T_i d_i \cdot c_t \\ &= v p_i e^{-bd_i} - p_i e^{-bd_i} d_i c_t \end{aligned}$$

We wish to find the gradient of NBU_i with respect to the location of the centre whose co-ordinates (X_j, Y_j) are treated as variables.

The gradient is given by the partial derivatives $\frac{d \text{NBU}_i}{dX_j}$ and $\frac{d \text{NBU}_i}{dY_j}$.

As v and c_t are constants

$$\frac{d \text{NBU}_i}{dX_j} = v \frac{dT_i}{dX_j} - c_t \frac{d(T_i d_i)}{dX_j}$$

Since an expression has already been derived for $\frac{dT_i}{dX_j}$ in Appendix 2, it only remains to obtain the derivative, $\frac{d(T_i d_i)}{dX_j}$.

Because the expression to be differentiated, $T_i d_i$, is a product we can use the general relation

$$\frac{d(uw)}{dx} = u \frac{dw}{dx} + w \frac{du}{dx}$$

Applying this general rule here

$$\frac{d(T_i d_i)}{dX_j} = T_i \frac{d(d_i)}{dX_j} + d_i \left(\frac{dT_i}{dX_j} \right)$$

Of the two terms on the left the second is relatively straightforward, consisting mainly of the derivative obtained in Appendix 2.

To obtain $\frac{d(d_i)}{dX_j}$ it is helpful to define $u = d_i^2$ as in the previous appendix. Then $d_i = u^{\frac{1}{2}}$ so d_i is a function of u which is

in turn a function of X_j as noted previously. Hence by the chain rule

$$\begin{aligned}\frac{d(d_i)}{dX_j} &= \frac{d(d_i)}{du} \cdot \frac{du}{dX_j} \\ &= \frac{1}{2} u^{-\frac{1}{2}} (-2x_i + 2X_j) \\ &= \frac{-2(x_i - X_j)}{2u^{\frac{1}{2}}} \\ \therefore \frac{d(d_i)}{dX_j} &= -\frac{x_i - X_j}{d_i}\end{aligned}$$

$$\text{Hence } \frac{d(T_i d_i)}{dX_j} = -T_i \frac{(x_i - X_j)}{d_i} + d_i \left(\frac{dT_i}{dX_j} \right)$$

$$\begin{aligned}\text{Thus } \frac{d \text{ NBU}_i}{dX_j} &= v \frac{dT_i}{dX_j} + c_t T_i \frac{(x_i - X_j)}{d_i} - d_i \left(\frac{dT_i}{dX_j} \right) \\ &= \frac{dT_i}{dX_j} (v - d_i) + c_t T_i \frac{(x_i - X_j)}{d_i} \\ &= bp_i (x_i - X_j) e^{-bd_i} d_i^{-1} (v - d_i) \\ &\quad + c_t p_i e^{-bd_i} \frac{(x_i - X_j)}{d_i} \quad (\text{from Appendix 2}) \\ \therefore \frac{d \text{ NBU}_i}{dX_j} &= p_i e^{-bd_i} (x_i - X_j) \left(v b d_i^{-1} - b + \frac{c_t}{d_i} \right)\end{aligned}$$

If we consider i as one of n points surrounding the facility at j then the total net balance of costs and benefits to all users of j is $\sum_{i=1}^n \text{NBU}_i$. The partial derivative of this total is

$$\frac{d(\sum_i \text{NBU}_i)}{dX_j} = \sum_i p_i e^{-bd_i} (x_i - X_j) (v b d_i^{-1} - b + \frac{c_t}{d_i})$$

When $\sum_i \text{NBU}_i$ is at a maximum this expression has the value 0 by

definition:

$$\therefore \sum_i x_i p_i e^{-bd_i} (vbd_i^{-1} - b + c_t d_i^{-1}) = X_j \sum_i p_i e^{-bd_i} (vbd_i^{-1} - b + c_t d_i^{-1})$$

$$\therefore X_j = \frac{\sum_i x_i p_i e^{-bd_i} (vbd_i^{-1} - b + c_t d_i^{-1})}{\sum_i p_i e^{-bd_i} (vbd_i^{-1} - b + c_t d_i^{-1})}$$

The corresponding equation for Y_j is given by symmetry. Both equations have to be solved iteratively since d_i is unknown. For the sake of simplicity c_t was given the value of 1 in all analysis involving NBU. With $c_t = 1$ this equation for X_j is identical to 6.12.

APPENDIX 4

The LOCHWISP algorithm

```

.....
. LOCHWISP .
.....

```

A program for examining the locations of central facilities on a plane.

Written in Edinburgh Fortran for the ICL 2970 by R.L.Hodgart, Dept. of Geog., Univ. of Edinburgh.

Data is read from two files :

- (a) parameters of problem, input format and location of population (channel 5);
- (b) m and data on facilities (channel 4).

In the first file the order of data line by line is

- (1) - parameters of problem as follows (one line)

N : no. of demand points (I4)

B : b, degree of elasticity (F4.0)

BSHAR : bs, defines catchment areas in spatial interaction model (F4.0)

S : s, covering radius (F4.0)

V : v, value of one trip (F4.0)

THIESS : YES gives Thiessen catchments (A4)
NO gives overlapping catchments

GOAL : 1 minimises travel cost ; 2 maximises use
3 maximises NBU (F4.0)

TOL : tolerance value used to end search (F4.0)

ITLIM : max. number of iterations allowed (I4)

NORL : 1 gives NORLOC type of grid (I4)

0 (i.e. otherwise) treats locations literally

STRAT : 1 allows each facility to move only once during an iteration (F4.0)

0 each facility moves till its displacement is less than TOL

PLOT : 1 plots searching path (F4.0)

0 (i.e. otherwise) no plot produced

- (2) - FMT, format to be used in reading data on population and facilities

- (3) etc. - x & y co-ords and pop'n of demand pts. (FMT)

In the second file the order line by line is

- (1) M : no. of centres, m (I4)

- (2) co-ords. of 1st centre, whether fixed (1.0) or moveable (0.0) and attractiveness (FMT)

- (3) same information for 2nd centre and so on line by line till mth centre

- (4) title of problem in one line (60A1)

- (5) MORE : YES means more problems are to be read in
NO terminates the program (A4)

- (6) title of new problem in one line (60A1)
 (7) MNEW : no. of new centres to be read in (I4)
 which must be $\leq m$ or $= m$ in this version
 (8) co-ords etc. of new centres ; if MNEW is 1 this is
 assumed to be for the mth centre ;
 if MNEW is 2 next two lines are locations etc.
 for (m-1)th & mth centres resp. and so on (FMT)
 (9) NO after co-ords etc. for last problem. (A4)

Use of MORE & MNEW is a device to make it easy to try a
 series of possible locations for additional facilities.
 Another available version of the main program provides more
 information about the impact of new centres on existing ones.

```
COMMON XP(400),YP(400),POP(400),DNEAR(400),XF(25),YF(25),FTAG(25)
COMMON NTRIB(25),HINT(25,400),AT(25),TR(25),CV(25),BUN(25)
COMMON M,N,B,S,V,THIESS,ALLOC(400),ATS,TRS,CVS,BUNS,DMAXS
COMMON CAPAC(25),CAPACS,CT,DMAX(25),EQIDIS(70,5),IEQSUM,TOL,ITER
COMMON MAXDIF,GOAL,ITLIM,ATSO,TRSO,BUNSO,GRD,STRAT,PLOT
COMMON XF1(25),YF1(25),TITLE(60),BSHAR,ATR(25)
REAL MAXDIF
INTEGER THIESS, YES/'YES'/'NO/'NO'/'',FMT(6)
INTEGER EQIDIS
INTEGER *2 TITLE
READ(5,51)N,B,BSHAR,S,V,THIESS,GOAL,TOL,ITLIM,NORL,STRAT,
* PLOT
READ(5,50)FMT
CT = 1.0
DO 4 I=1,N
4 READ(5,FMT) XF(I),YP(I),POP(I)
  READ(4,53)M
  DO 3 J=1,M
  READ(4,FMT) XF(J), YF(J),FTAG(J),ATR(J)
  XF1(J)=XF(J)
  YF1(J)=YF(J)
3 CONTINUE
  READ(4,49) TITLE
  WRITE(6,62)
  IF(THIESS.EQ.YES) WRITE (6,63)
  IF(THIESS.EQ.NO) WRITE(6,64)
  GRD=0.0
  IF(NORL.EQ.1)GRD=0.5
  IF(NORL.EQ.1) WRITE(6,72)
  IF(PLOT.EQ.1) WRITE(6,661)
  IF(GOAL.EQ.1)WRITE(6,66)
  IF(GOAL.EQ.2)WRITE(6,67)B
  IF(GOAL.EQ.3) WRITE(6,671)B,V
  WRITE(6,65) M,N,B,BSHAR,S,V,TOL
  WRITE(6,54)ITLIM
  WRITE(6,55)
  DO 5 J=1,M
  WRITE(6,56) J,FTAG(J),ATR(J)
5 CONTINUE
90 WRITE(6,48) TITLE
  IEQSUM = 0
  MAXDIF=0.0
```

```

      ATSO=1.0E10
      TRSO=0.0
      BUNSO=0.0
      ITER=1
99  WRITE(6,68) ITER
      IF(THIESS.EQ.YES) CALL CATCH
      CALL EVALOC
      IF(ITER.GT.ITLIM) GO TO 6
      GO TO 99
6   IF(IEQSUM.GT.0)CALL ADJUST
      READ(4,57) MORE
      IF(MORE.NE.YES) GO TO 101
      READ(4,49)TITLE
      READ(4,58) MNEW
      DO 7 II=1,MNEW
      KK=M-MNEW+II
      READ(4,FMT) XF(KK),YF(KK),FTAG(KK),ATR(KK)
      XF1(KK)=XF(KK)
      YF1(KK)=YF(KK)
7   CONTINUE
      GO TO 90
48  FORMAT(/4X,'TITLE : ' 60A1 )
49  FORMAT(60A1)
50  FORMAT(10X,6A4)
51  FORMAT(I4,4F4.0,A4,2F4.0,2I4,2F4.0)
52  FORMAT(10X,3F6.0)
53  FORMAT(I4)
54  FORMAT(/3X' MAXIMUM ITERATIONS ='I4)
55  FORMAT(/10X, ' FAC   FIXED  ATTRAC ')
56  FORMAT( 10X, I4,2F8.1)
57  FORMAT(A4)
58  FORMAT(I4)
62  FORMAT(/4X,'LOCATION OF CENTRES BY LOCHWISP :PLANE GEOMETRY' )
63  FORMAT(/,4X,'THIESSEN CATCHMENTS' )
64  FORMAT(/,4X,'OVERLAPPING CATCHMENTS' )
65  FORMAT(/4X'M ='I4,4X'N ='I4,4X'B ='F7.3,5X'BS VAL FOR ALLOC MODEL
      * ='F7.3/3X, 'S =' F6.2,3X 'V =' F6.2,4X ' TOL =' ,F6.3 )
661 FORMAT( 3X,' PLOTTING ROUTINE INVOKED' )
66  FORMAT(/4X'SEARCH : MINIMISING TRAVEL COST' )
67  FORMAT(/4X'SEARCH : MAXIMISING USE -ELASTIC DEMAND'/13X'B ='F7.3)
671 FORMAT (/4X'SEARCH : MAXIMISING NET BENEFIT'/13X'B ='F6.3
      *,4X,'V =' F6.1)
68  FORMAT(/' ** ITERATION ** ',I4)
72  FORMAT(/ 4X'NORLOC TYPE OF GRID USED IN COMPUTING DISTANCES')
101 STOP
      END

SUBROUTINE CATCH
COMMON XP(400),YP(400),PDF(400),DNEAR(400),XF(25),YF(25),FTAG(25)
COMMON NTRIB(25),HINT(25,400),AT(25),TR(25),CV(25),BUN(25)
COMMON M,N,B,S,V,THIESS,ALLOC(400),ATS,TRS,CVS,BUNS,DMAXS
COMMON CAPAC(25),CAPACS,CT,DMAX(25),EQIDIS(70,5),IEQSUM,TOL,ITER
COMMON MAXDIF,GOAL,ITLIM,ATSO,TRSO,BUNSO,GRD,STRAT,PLOT
COMMON XF1(25),YF1(25),TITLE(60),BSHAR,ATR(25)
REAL MAXDIF
INTEGER THIESS,YES/'YES' /,NO/'NO'  /
INTEGER*2 TITLE
INTEGER EQIDIS

```

```

IEQSUM=0.0
K=1
NEARF=1
DO 12 J=1,M
12 NTRIB(J) = 0.0
DO 13 I =1,N
DMIN = 9999999999.9
DO 14 J=1,M
DC =( GRD+ XP(I)-XF(J) )**2 + ( GRD+ YP(I)-YF(J) )**2
IF(DC.NE.DMIN) GO TO 33
IEQSUM = IEQSUM +1
EQIDIS(K,1) = I
EQIDIS(K,2) = POP(I)
EQIDIS(K,3) = NEARF
EQIDIS(K,4) = J
K = K+1
33 IF (DC.GE.DMIN) GO TO 14
DMIN = DC
NEARF = J
14 CONTINUE
ALLOC(I) = NEARF
NTRIB(NEARF) = NTRIB(NEARF) + 1
L = NTRIB(NEARF)
HINT(NEARF,L) = I
DMIN = SQRT(DMIN)
DNEAR(I) = DMIN
13 CONTINUE
WRITE(6,70)
DO 15 J =1,M
15 WRITE(6,71)J, NTRIB(J)
IF(IEQSUM.EQ.0) GO TO 79
WRITE(6,73)
DO 18 K=1,IEQSUM
18 WRITE(6,74) (EQIDIS(K,K2),K2=1,4)
70 FORMAT( /,' CENTRE NO OF TRIB POINTS' )
71 FORMAT( 15,5X,I10 )
73 FORMAT(/'DEMAND POINTS WITH EQUIDISTANT FACILITIES'// 'DEM PT PO
*PULN FAC1 FAC2')
74 FORMAT(2X,I4,I10,2I6)
79 RETURN
END

SUBROUTINE EVALOC
COMMON XP(400),YP(400),POP(400),DNEAR(400),XF(25),YF(25),FTAG(25)
COMMON NTRIB(25),HINT(25,400),AT(25),TR(25),CV(25),BUN(25)
COMMON M,N,B,S,V,THIESS,ALLOC(400),ATS,TRS,CVS,BUNS,DMAXS
COMMON CAPAC(25),CAPACS,CT,DMAX(25),EQIDIS(70,5),IEQSUM,TOL,ITER
COMMON MAXDIF,GOAL,ITLIM,ATSO,TRSO,BUNSO,GRD,STRAT,PLOT
COMMON XF1(25), YF1(25),TITLE(60),BSHAR,ATR(25)
INTEGER THIESS,YES/'YES'/'NO'/'NO'/'NO',EQIDIS
INTEGER*2 TITLE
REAL ATC(25),NUMX(25),NUMY(25),DEN(25),MAXDIF,POP2(400)
REAL XFPL(25,12),YFPL(25,12),XFIX(25),YFIX(25),XF0(25),YF0(25)
DIMENSION XFOLD2(25),YFOLD2(25)
IF(ITER.EQ.1)WRITE(6,24)
IF(ITER.EQ.1)IFIX=1
MAXDIF=0.0

```

```

PROB=0.0
TRS =0.0
ATCS = 0.0
ATS =0.0
BUNS =0.0
CVS = 0.0
DMAXS = 0.0
CAPACS =0.0
WRITE(6,25)
DO 18 J=1,M
  XFOLD2(J)=XF(J)
18 YFOLD2(J)=YF(J)
  DO 21 J=1,M
    ITERJ=0
20 NUMX(J)=0.0
    NUMY(J) = 0.0
    DEN(J) =0.0
    TR(J) = 0.0
    ATC(J) = 0.0
    AT(J) = 0.0
    BUN(J) = 0.0
    CV(J) = 0.0
    CAPAC(J) = 0.0
    DMAX(J) = 0.0
    XFOLD=XF(J)
    YFOLD=YF(J)
    IF(THIESS.EQ.YES) L=NTRIB(J)
    IF(THIESS.EQ.NO) L = N
    DO 22 K =1,L
      IF(THIESS.EQ.YES) LL= HINT(J,K)
      IF(THIESS.EQ.NO) LL= K
      IF(ITERJ.EQ.0.AND.THIESS.EQ.YES) DC = DNEAR(LL)
      IF(ITERJ.NE.0) DC=SQRT((GRD+XF(LL)-XFOLD)**2+(GRD+YP(LL)-YFOLD)**2)
      IF(THIESS.EQ.NO.AND.ITERJ.EQ.0) DC=SQRT((GRD+XF(K)-XFOLD)**2+(GRD
* +YP(K)-YFOLD)**2)
      IF(THIESS.EQ.YES) GO TO 78
      POP2(LL) = POP(LL)
      CALL SHARE( DC,M,LL,POP,BSHAR,XP,YP,XF,YF,XFOLD2,YFOLD2,ATR,J)
78 TRZ = POP(LL)*EXP(-B*DC)
      ATZ = TRZ*DC
      TR(J)= TR(J) +TRZ
      ATC(J) = ATC(J) + ATZ
      AT(J) = AT(J) + POP(LL)*DC
      IF(V.LT.CT*DC) PROB=1.0
      IF(V.GE.CT*DC) BUN(J)=BUN(J)+ V*TRZ - CT* ATZ
      IF(DC.LE.S) CV(J) = CV(J) +POP(LL)
      CAPAC(J) = CAPAC(J) + POP(LL)
      IF(DC.GT.DMAX(J)) DMAX(J) = DC
      IF(FTAG(J).EQ.1) GO TO 19
      IF(GOAL.EQ.2) GO TO 70
      IF(GOAL.EQ.3) GO TO 75
      DEN(J) = DEN(J) +POP(LL)/DC
      NUMX(J) = NUMX(J) + POP(LL)*XF(LL)/DC
      NUMY(J) = NUMY(J) + POP(LL)*YP(LL)/DC
      GO TO 19
70 DEN(J) = DEN(J)+TRZ/DC
      NUMX(J) = NUMX(J) + XF(LL)*TRZ/DC
      NUMY(J) = NUMY(J) + YF(LL)*TRZ/DC

```

```

      GO TO 19
75  VBD=V*B/DC+1/DC-B
      DEN(J)=DEN(J)+TRZ*VBD
      NUMX(J)=NUMX(J)+XP(LL)*TRZ*VBD
      NUMY(J)=NUMY(J)+YP(LL)*TRZ*VBD
19  IF(THIESS.EQ.NO)POP(LL)=POP2(LL)
22  CONTINUE
      ITERJ=ITERJ+1
      XF(J)=XFOLD
      YF(J)=YFOLD
      IF(FTAG(J).EQ.1) GO TO 33
      XF(J)=NUMX(J)/DEN(J)
      YF(J)=NUMY(J)/DEN(J)
      DIFD=SQRT((XFOLD-XF(J))**2+(YFOLD-YF(J))**2)
      IF(ITER.EQ.1)GO TO 34
      IF(STRAT.EQ.1) GO TO 34
      IF(DIFD.GT.TOL) GO TO 20
34  IF(DIFD.GT.MAXDIF) MAXDIF=DIFD
33  IF(PLOT.NE.1)GO TO 35
      XFPL(J,ITER)=XFOLD
      YFPL(J,ITER)=YFOLD
      IF(FTAG(J).NE.1.OR.ITER.NE.1) GO TO 35
      XFIX(IFIX)=XFOLD
      YFIX(IFIX)=YFOLD
      IFIX=IFIX+1
35  TRS = TRS +TR(J)
      ATCS = ATCS + ATC(J)
      ATS = ATS + AT(J)
      BUNS = BUNS + BUN(J)
      CVS = CVS +CV(J)
      CAPACS = CAPACS + CAPAC(J)
      IF(DMAX(J).GT.DMAXS) DMAXS =DMAX(J)
      WRITE(6,26)J,XFOLD,YFOLD,TR(J),ATC(J),AT(J),BUN(J),CV(J)
      *,CAPAC(J),DMAX(J),ITERJ
21  CONTINUE
      WRITE(6,27) ITER,TRS,ATCS,ATS,BUNS,CVS,CAPACS,DMAXS
      IF(ITER.NE.1)GO TO 36
      TRS1=TRS
      ATCS1=ATCS
      ATS1=ATS
      BUNS1=BUNS
      CVS1=CVS
      CAPAS1 =CAPACS
      DMAXS1=DMAXS
36  IF(ITER.LT.ITLIM) GO TO 37
      TRS1=(TRS-TRS1)/TRS1*100
      ATCS1=(ATCS-ATCS1)/ATCS1*100
      ATS1=(ATS-ATS1)/ATS1*100
      BUNS1=(BUNS-BUNS1)/BUNS1*100
      CVS1=(CVS-CVS1)/CVS1*100
      DMAXS1=(DMAXS-DMAXS1)/DMAXS1*100
      WRITE(6,30) TRS1,ATCS1,ATS1,BUNS1,CVS1,DMAXS1
      IF(ITER.LT.ITLIM)GO TO 37
      IF(ITER.EQ.1)GO TO 37
      WRITE(6,41)

```

```

DO 42 J=1,M
  DMOV= SQRT( (XF1(J)-XF0(J))**2+(YF1(J)-YF0(J))**2 )
42 WRITE(6,43) J,DMOV
37 IF(PROB.EQ.1) WRITE(6,28)
  IF(GOAL.EQ.1.AND.ATS.GT.ATSO)ITER=ITLIM+1
  IF(GOAL.EQ.2.AND.TRS.LT.TRSO)ITER=ITLIM+1
  IF(GOAL.EQ.3.AND.BUNS.LT.BUNSO)ITER=ITLIM+1
  IF(ITER.GT.ITLIM)WRITE(6,31)
  ATSO=ATS
  TRSO=TRS
  BUNSO=BUNS
  ITER= ITER+1
  WRITE(6,29) MAXDIF
  NFIX=IFIX-1
  IF(PLOT.EQ.1.AND.ITER.GT.ITLIM) CALL JOHN(XFPL,YFPL,M,ITLIM,XFIX,
& YFIX,NFIX,B,V,GOAL,THIESS,TITLE,ATS1,TRS1,BUNS1,ITER,BSHAR)
24 FORMAT(/ 'EVALUATION OF INITIAL POSITIONS')
25 FORMAT(/ 'CEN   XF   YF   TRIPS   ELA TC   INEL TC   NET BENU
*POP COVD   CAPAC MAXD ITERJ' )
26 FORMAT(I2,2F6.2,6F10.1,F6.1,I5)
27 FORMAT(/'ITR',I3,' SYSTEM ',6F10.1,F6.1 /)
28 FORMAT(/ 4X,'PROBLEM OF V LT TRAV COST HAS OCCURED' )
29 FORMAT(/' LARGEST DISPLACEMENT (MAXDIF) =' F9.3 )
30 FORMAT(' %CHANGE FROM'/ ' INITIAL LOCN',F9.2,4F10.2,'   NOT AP '
* ,2F6.2)
31 FORMAT(/' DETERIORATION IN ONE OF THE OBJECTIVE FUNCTIONS' )
41 FORMAT(/,1X ' DISTANCES MOVED FROM FIRST POSITIONS (IN 500M UNITS
&)' )
43 FORMAT( 12X, I3,5X,F10.2 )
  RETURN
END

SUBROUTINE SHARE(DC,M,LL,POP,BSHAR,XP,YP,XF,YF,XFOLD2,YFOLD2,ATR
* ,J)
  REAL POP(400),XP(400),YP(400),XF(25),YF(25),XFOLD2(25),YFOLD2(25),
* ATR(25)
  TOTATT=0.0
  DO 91 JJ=1,M
    DISLLJ = SQRT( (XP(LL)-XFOLD2(JJ))**2 +(YP(LL)-YFOLD2(JJ))**2 )
    IF(JJ.EQ.J)DCJ=DISLLJ
    TOTATT=TOTATT+ATR(JJ)*EXP(-BSHAR*DISLLJ)
91  CONTINUE
  PROP=(ATR(J)*EXP(-BSHAR*DCJ))/TOTATT
  POP(LL) = POP(LL)*PROP
  RETURN
END

```

C
 C The example below involves the location of 1 additional centre,
 C given two existing ones (fixed), to serve 10 demand points.
 C A second starting location for the extra centre is tested
 C after the first.

C
 C (a) Input on channel 5

C
 10 .25 .25 5.0 5.YES 1.0 .20 3 0 0.0 0.0
 FMT (SX,2FS.0,F11.1,F5.2)
 12.5 21.5 48.0
 21.5 19.5 4339.0
 17.5 17.5 486.0
 26.5 16.5 800.0
 12.5 14.5 372.0
 20.5 13.5 1549.0
 29.5 12.5 3079.0
 14.5 10.5 2847.0
 24.5 9.5 322.0
 7.5 7.5 949.0

C
 C (b) Input on channel 4
 C

3
 16.2 12.6 1.0 1.00
 20.2 13.4 1.0 1.22
 21.6 11.8 0.0 5.60
 Start at Park Road
 YES
 Start at Roseburn
 1
 13.2 13.3 0.0 5.6
 NO

APPENDIX 5The effective extent of demand cones and catchment areas

In trying to understand what a given value of b really means in terms of how far the associated demand cone actually extends around a centre, it is helpful to consider initially a uniform landscape occupied by one facility. It is possible to calculate how much demand will come to this centre from within a given radius by integrating the demand cone up to that limit; this involves evaluating equation 3.13 for the appropriate value of a . Similarly, the total demand under the cone can be estimated accurately by evaluating the same equation with a very high value of a . It is then possible to express the former demand as a percentage of the total. Such values have been tabulated below to illustrate the implications of various b -values. Thus, if b was .125, less than 36% of the centre's demand would come from within a radius of 10 units of distance, whereas, if b was .25 or .50, the corresponding figures would be 71.3% and 96.0%. In these terms, there is therefore a substantial difference between b -values of .125 and .25.

(see table overleaf)

<u>Radius</u>	<u>Percentage of all trips to the centre which come from within the specified radius for the given value of b or b_s</u>		
	<u>b or b_s = .125</u>	<u>b or b_s = .25</u>	<u>b or b_s = .50</u>
1	0.7	2.7	9.0
2	2.7	9.0	26.4
3	5.5	17.3	44.2
4	9.0	26.4	59.4
5	13.0	35.5	71.3
6	17.3	44.2	80.1
7	21.8	52.2	86.4
8	26.4	59.4	90.8
9	31.0	65.8	93.9
10	35.5	71.3	96.0
15	55.9	88.8	99.5
20	71.3	96.0	100.0
30	88.8	99.5	100.0

When several centres are located on the plane the tabulated figures still apply to each, whether they are considered as b-values used to define the degree of elasticity or as b_s-values used only to define the catchments of the centres. Of course, in real situations these figures no longer hold, partly because the density of population varies. However, they can still be taken as an indication of the tendency implicit in particular values.

APPENDIX 6List of mathematical symbols used in the text

In choosing symbols there has been a conflict between internal consistency and consistency with the notation used in parts of the literature or that in general use. Partly for this reason and partly for convenience some symbols have been given a local definition which is not consistent with their use in the rest of the thesis. For instance, c is used to denote the constant term in an integration in parts of Chapter 3, but is used in Chapter 4 for the cost of constructing one facility; v is used on page 88 to denote any function but thereafter denotes the value of a trip.

(a) Lower case symbols

- | | |
|----------------------|--|
| a | radius of circle inscribed within a hexagonal catchment area |
| a' | distance beyond which the cost of a trip to centre j exceeds its value |
| $a_1, a_2 \dots a_m$ | radii of circles inscribed within hexagonal catchments produced with 1, 2 ... or m facilities in existence. |
| a_{ij} | binary coefficient describing either pattern of cover from facility j or assignment of demand points to facilities according to the problem being considered: when $a_{ij} = 1$ then i is covered by j or i is assigned to j ; when $a_{ij} = 0$ i is not covered by j or is not assigned there. |
| b | exponent measuring decrease in use or demand with distance i.e. spatial elasticity of demand |
| b_s | exponent used to define catchment areas when these overlap; in a sense it is the rate at which the proportion of users in i allocated to j tends to decline with distance. |
| b_j | cost of expanding centre j by one unit of population or demand |
| c | constant term of an integration (Chapter 3) |
| c | cost of construction of one centre (Chapter 4 and elsewhere) |

c_t	cost of travel per km
d_{ij}	shortest distance between i and j on a plane or through a network
f	frequency of trips per head of population to a facility from a given area during a specified period
f_0	constant defined by value of f for a point at zero distance from a facility
f_j	cost of opening facility j
g, h	functions
i	a particular demand point or population cell
j	a particular facility
k	the number of additional facilities to be allocated (Chapter 5)
k	a particular sub-group within a population cell (Chapter 6)
k_1	a constant $= 2\pi p$
k_{AT}	a constant $= .7020 \ 2\sqrt{3} \ a_1^3 p$
k_2	a constant used to define elastic demand (page 74)
k_3	a constant $= k_2 k_1$
k_4	a constant $= \frac{k_3}{2-b} a_1^{2-b}$
ℓ	the number of sub-groups in each population cell (Chapter 6)
m	the number of facilities
n	the number of demand points or cells of population
p	density of population on an isotropic plane
p_i	demand or population at i
p_{ik}	demand from or population in sub-group k at cell i
q	a minimum standard of access deemed to be desirable and defined in terms of distance
r	distance of annulus from centre of circular catchment
r_{max}	radius of circle circumscribed around hexagonal catchment
r_{min}	radius of circle inscribed within hexagonal catchment
s	radius of cover provided by a facility at j

u	a function
v	the monetary value to one individual of using a facility once; but also used locally on page 88 to denote any function
w	a function
w_j	the attractiveness of facility j
w_k	the average demand per head of sub-group k
x_i	the x co-ordinate of i
y_i	the y co-ordinate of i
z	used to denote various objective functions

(b) Upper case symbols and abbreviations

AT	aggregate travel cost in km
AT_i	travel cost in km emanating from i
AT_j	aggregate travel cost in km within catchment of j
AT_m	aggregate travel cost in km to all m facilities
BU	benefit to whole population of users of $\sum_j T_j$ trips = $v \sum_j T_j$
H_i	aggregate travel cost in km of any approximate solution i produced by one trial of a heuristic
IA	inequality in access i.e. difference between nearest and furthest users
MT	mean travel distance
MT_j	mean travel distance within catchment of j
NBU	net benefit to users, i.e. difference between benefits of use (BU) and costs of travel (TC) over the whole system of facilities
NSB	net social benefit, i.e. difference between net benefit to users, NBU, and total cost of supply, SC.
O_i	aggregate travel cost in km of global optimum for problem i
PC	population covered by all j facilities
SC	total cost of supplying m facilities

T_{ij}	number of trips from cell i to facility j
T_i	number of trips emanating from i
TC	total travel cost over the whole system in £ ($= AT.c_t$)
TC_m	travel cost to all m facilities in £
TS_m	total saving in travel costs resulting from construction of m facilities as compared to situation with 1 facility

REPRINTED FROM

Progress in Human Geography

Volume 2, Number 1, 1978



Edward Arnold

Printed in Great Britain at the Alden Press, Oxford

Optimizing access to public services:

a review of problems, models and methods of locating central facilities

by R. L. Hodgart

It must also be confessed, that, wherever we depart from this equality, we rob the poor of more satisfaction than we add to the rich, and that the slight gratification of a frivolous variety, in one individual, frequently costs more than bread to many families, and even provinces (D. Hume, *An enquiry concerning the principles of morals*).

Over a decade and a half ago Bunge (1962, 196) focused attention on the problem of placing 'interacting objects as near to each other as possible' in geographical space, suggesting it was one of geography's central questions. Since then substantial progress has been made on one of such problems formerly thought to be intractable: the problem of locating a given number of facilities, such as clinics or public libraries, so that the population concerned enjoys the best possible geographical access to the service. Here the interacting objects are the centres of supply treated as points, and the units of demand or need, usually treated as areas, grid squares or nodes.

Impressive progress has been made in formulating and solving the problem mathematically by a variety of optimization methods (ReVelle *et al.*, 1970; Scott, 1971; Törnqvist *et al.*, 1971; Rushton *et al.*, 1973). Numerous papers in the various journals of mathematical geography, operations research, management science, regional science and planning testify to the growing flexibility and efficiency of approaches now available. More recently there has been an increasing awareness that such mathematical approaches are not free of social value judgements (Doherty, 1973; Morrill, 1974; Dear, 1974a; McAllister, 1976) and that models should be formulated to optimize goals of equity as well as efficiency.

Good introductions to this work can now be found in a number of geographical texts (Abler *et al.*, 1971; Massam, 1972; Taylor, 1977). Massam (1974) has outlined various ways of solving one form of the problem and in the later work (1975) discussed the literature bearing on the problem from many disciplines. An extremely valuable bibliography on location/allocation systems by Lea (1973) provides summaries and comments on over 600 items, including Soviet work. So far, however, the geographical literature does not appear to have produced a critical assessment of the assumptions and social objectives of such models within the wider context of spatial interaction and location theory. The main aim of the present article will be to outline a basis for such an assessment.

I Public facilities and the political planning process

A decision to locate any public facility is essentially a decision to distribute certain benefits and costs among different groups of people. The benefits to adjacent communities may include improved access to health care or the enhanced amenity to a neighbourhood if a park is sited nearby. The costs may include loss of local amenity due to a new sewage plant or a new road. Very often these benefits and losses will be related in some consistent way to proximity, which may allow the unpriced gains and losses involved to be treated as a function of distance from the facility (Harvey, 1973; Austin, 1974).

On this basis facilities can be classified broadly into three types:

- 1 Central facilities to which users must travel to obtain the service and which are used by most of the population are generally regarded favorably by local residents. They can also be called desirable facilities. It is usually assumed that they have no objectionable attributes and that the social benefits of a potential site can be measured by how much it reduces the cost or inconvenience of users' travel. Where the service is delivered from a central point (e.g. fire services) the agency's reduction in transport cost or time can be used in the same way.
- 2 Noxious facilities, such as garbage incinerators, may be needed by the area as a whole, but they impose costs such as increased noise or pollution on nearby residents. Naturally, plans to construct such facilities are frequently opposed by local groups. The loss of amenity may again be treated partly as a function of proximity.
- 3 Hybrid facilities confer a mixture of costs and benefits on local groups, the net balance of the two depending perhaps on social group and distance. A noisy discotheque which disturbs residents very close to it may be viewed more neutrally further away; and may be regarded as a valuable amenity by local teenagers.

Austin (1974) outlines a framework which encompasses the impacts of all three categories of facility on the local environment. More work is needed on the attitude of individuals to various kinds of facilities at different distances to provide evidence for classifying particular facilities (see Wolpert *et al.*, 1972).

The present study is mainly concerned with facilities of the first type where the goal is to maximize accessibility. A desirable goal for noxious facilities on the other hand may be to minimize accessibility to population, subject to site constraints. Austin *et al.* (1970) and Mumphrey and Wolpert (1973) have done much to clarify conceptually the complex questions of equity, compensation, and conflict resolution which lie at the heart of the noxious and hybrid facility problems, but it is inherently difficult to make optimizing models for these problems operational.

The work of Wolpert and that of Harvey (1973) has, however, drawn attention to the fact that the success of any group in western society in obtaining a just share of the public goods and benefits which are allocated

spatially through the political/planning process is likely to depend as much on class and political bargaining strength as on need. Noxious facilities are often sited in those parts of cities where people are least able to oppose the plan (Wolpert *et al.*, 1975). Various workers have documented inequalities in the allocation of medical services at a regional scale in Britain and the USA (Coates and Rawstron, 1971; Davies, 1968; de Vise, 1973). Galbraith (1974) and Navarro (1976) reach similar general conclusions, from different ideological perspectives. In a lighter vein, Pahl (1975, 304-6) recounts how similar spatial inequalities might arise in capitalist and socialist societies. Węclawowicz (1975) shows that social groups are less segregated in Warsaw than in western cities. How this affects the distribution of services in Polish cities cannot yet be assessed but a study of Poznań by Polarczyk (1976) is interesting in this context.

The final decisions about the allocation of all public resources are, of course, political. Given these realities what useful role can locational optimizing models play? First, by showing that better solutions exist, they give the community group a stronger case against inefficient and inequitable proposals. Second, they provide a means for weighing up different combinations of equity and efficiency as McGrew and Monroe (1975) show.

There are several further reasons why models of facility location may be of interest to geographers. They allow the traditional industrial location problem of Weber to be solved rapidly under various transport conditions on a plane (Cooper, 1968) and in a network (Kuehn and Hamburger, 1963; Hakimi, 1964). They may allow the Löschian location problem to be explored in less restrictive environments than the traditional isotropic plane. Finally, location-allocation procedures have certain geometrical affinities with electoral districting, taxonomic description and regionalization problems (Scott, 1970).

Historically, most models for facility location were formulated as private sector models for locating manufacturing plants, warehouses or depots, as Lea's bibliography (1973) shows. Essentially these models minimized the cost of transporting fixed amounts of inputs from their origins and/or fixed amounts of output to a market. Thus inputs and demand were assumed to be known in advance; they were not dependent on the location itself; consequently, competitive strategies from other firms could be ignored. If demand, expressed as a number of trips from the population units, does not depend on location of the facilities, the absence of competition makes it possible for public agencies also to locate so that the total transport costs of users are minimized. Hotelling (1929) illustrates the difference between this social or welfare optimum and the competitive equilibrium of private retailers. The latter is suboptimal under Hotelling's restrictive conditions. These arguments assume that the public agency will locate optimally, which may not happen owing to the absence of direct pressure from demand and the vagaries of the political/planning process. In addition, given the dynamic character of population distribution and transport technology, any static solution will only be optimal in the short

term. Scott (1975) illustrates how a dynamic programming framework can be used to explore the differences between welfare and competitive optima and to locate facilities under various time horizons. One obvious difficulty here is that the distribution of population is not easy to predict.

The problem of locating public services often involves essential services to which everyone must go, unlike the depot location problem. Equity of access is therefore an additional goal for public services (Dear, 1974b): a slightly less efficient but more equitable solution may be preferred to the global optimum.

II Forms of the problem and basic assumptions

Three basic elements can be varied in designing a system of facilities, namely their number, location and capacity. For many public services, as Toregas and ReVelle (1972) have noted, constraints on capacity are inappropriate unless queueing results in variable waiting times or people are assigned by administrative fiat to particular facilities as in electoral districting (Wagner and Falkson, 1975). In a sense the capacity of many facilities can be regarded as dependent on and therefore subordinate to their locations. In order to concentrate on location, discussion of capacity will be postponed until later, i.e. we will be concerned with the 'unconstrained problem'.

Three spatial forms of the problem can be identified:

- 1 assign m facilities freely, i.e. assume no facilities already exist in the area (the general problem);
- 2 locate k additional facilities, taking the existing centres into account (the additional or incremental facility problem);
- 3 given m existing centres, reorganize the system by closing any badly located centres and allowing a certain number to be opened (the reorganization problem).

Most writers have concentrated on the first problem, perhaps because the second is easier computationally; the third has received very little attention but can be solved by adapting methods for the first. Except in the rare case of an entirely new service or the unlikely event of all existing capital being written off, the second problem is much more likely to occur in reality than the first. This is worth emphasizing because it has one idiosyncrasy which will be discussed later.

Problems can also be classified according to whether the spatial framework of a network or a plane is used. Essentially the spatial framework defines the rules for measuring distance: shortest path in the network; Euclidean or city block distance on a plane.

A crucial distinction, and one which perhaps has not received sufficient attention, is that between elastic and inelastic demand models. In elastic demand models the individual's demand or use, expressed in trips to a centre, depends on the price of obtaining the service. Where public services

allow users to enter without a charge (e.g. public libraries, museums, parks), the price of the service can be taken simply as the travel cost. Where there is a standard entry charge (e.g. for many swimming pools in a city) the spatial variation in price is essentially a function of the variable travel cost, the other component of price being constant. The demand from population unit i will then depend on its level of access and therefore on the location of service centres (Smolensky *et al.*, 1970).

When the entry charge is zero, a corollary of elastic demand is that, as the number of facilities increases and the general level of access consequently improves, price is reduced and utilization is thereby increased (Figure 1). Thus, in what is now a seminal paper on the theory of facility location, Teitz argues that 'the system is in a curious position of being able to generate demand by organizing itself appropriately' (Teitz, 1968, 44). A further implication of elastic demand is that, other things being equal, supply should be organized in smaller, more closely spaced facilities than would be the case for a service with the same supply costs but inelastic demand. For the same supply cost more utilization and more revenue, where relevant, will be generated. Hence, the more elastic the demand with respect to travel cost, the stronger is the argument for smaller and more numerous facilities.

Despite the preceding arguments, inelastic models have so far dominated the field, with a few interesting exceptions (Abernathy and Hershey, 1971; Wagner and Falkson, 1975). Existing texts, such as Massam (1974; 1975) and Scott (1971), concentrate entirely on models assuming inelastic demand, which means that the demand of any unit i is taken as completely independent of where the facilities are located. Such models are sometimes called 'fixed requirement' models. It can be shown that increasing m brings successively smaller reductions in aggregate travel costs when demand is inelastic (Figure 1).

It is important not to confuse elastic demand with the well-known distance decay effect around individual facilities. A simple example may help to clarify this distinction. Consider a certain service with inelastic

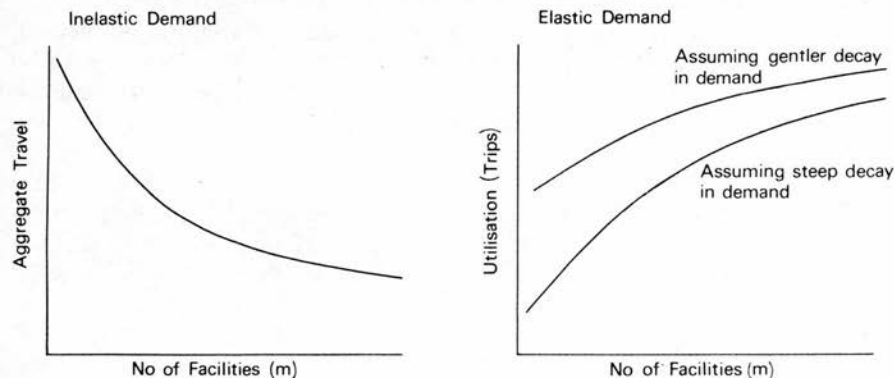


Figure 1 Benefit derived from increasing the number of facilities under conditions of (a) inelastic demand, (b) elastic demand.

demand which is met by four facilities in a certain city. The number of trips from particular population subareas may then vary with such factors as age, income and social composition but will be invariant with respect to the location of the centres. Suppose a certain proportion of users from any subarea, i , go to their nearest centre, and successively smaller proportions go to the second, third and fourth nearest facilities. Around any facility, j , the density of trips (i.e. the proportion of users who prefer to use j) will be higher near the facility and much lower for areas further away where, for instance, j may only be the fourth nearest centre. A study of the catchment area of one of the facilities will then reveal a distance decay effect, though demand is inelastic spatially. In short, if demand is inelastic, distance decay describes how distance affects choice of centre; whereas with elastic demand models, distance decay firstly describes how demand falls with distance. In the latter case a second decay effect may subsequently be used to describe the choice of centres.

An example of a price sensitive model is given by Abernathy and Hershey (1971). In their model the number of trips from an area depends on city block distance to the nearest facility; these trips are then allocated to centres by a probabilistic function based on distance. The population units in their model can be disaggregated into social subgroups, each with different distance parameters for the decay in both demand and preference. Essentially this model can be viewed as an elastic spatial interaction model. Since some of the spatial interaction models formulated by Wilson (1974) could also be applied in this field it is convenient to note here that they mostly fall into the inelastic category. The classical models of Lösch and Christaller are, of course, based on elastic demand.

Although several studies report distance decay effects around supply points (Berry, 1967; Taylor, 1971), their design often does not permit an assessment of the extent to which this represents a fall in demand as opposed to a decline in preference. This question can only be approached by comparing the patterns of use of individuals with high and low access through a household survey or through a survey of users at all facilities in an area. The results of a study of the latter kind by Weiss and Greenlick (1970) suggest that, for medical services, distance mainly affects the choice of facility rather than frequency of use. The evidence of an exploratory study of the use of public swimming pools in Edinburgh (Currie, 1977) tentatively supports the argument that demand for this service is fairly elastic with respect to access.

The pattern of social class segregation within cities can further complicate the interpretation of survey results. In a study of children's dental health in Newcastle-upon-Tyne by Bradley *et al.* (1976), it was found that the percentage of children needing treatment correlated as strongly with accessibility as with social class at the scale of school catchment areas. This finding can be explained in two ways. The first, implied by the authors, is that accessibility affects the frequency of visits to a dentist and therefore the condition of teeth. But an alternative partial explanation is that dentists tend to locate in disproportionate numbers in middle class parts of the city

and so the higher income groups, who encourage their children to take better care of their teeth anyway, are generally more accessible to dentists. The correlation with accessibility is then partly fortuitous. A more finely disaggregated survey might help to resolve the problem of interpretation. In either case, however, there is an argument for encouraging a pattern of dental care which is less concentrated spatially than is normally the case in British cities: firstly, to improve utilization of the service and thereby dental health: secondly, to improve equity of access to an essential service.

The nature of demand for some services, however, seems relatively clear cut. For obvious reasons it seems sensible to assume that demand for fire services, for secondary education and for the most essential medical services is strongly inelastic. In such cases it would be just as appropriate to speak of 'need' as of demand. The concept of 'supply-led demand' (Coppock and Duffield, 1975) often quoted in relation to recreation services is equivalent to postulating elastic demand. Clearly more research is needed to determine how elastic is the demand for those services which are not so easy to classify. The conclusions of such studies will have a significant bearing on what kind of objective function is appropriate for a particular service. Of course in such studies there is always a danger of confusing demand, which is partly a function of income, with need, which is not dependent on income but harder to define.

III Objectives

The outcome of an optimizing model depends on its objectives and constraints. As an interesting paper by Eilon (1972) shows, optimizing models are more flexible than their use hitherto in geography would indicate because goals and constraints can often be interchanged. In order to clarify the various goals, their implications and bias, it is convenient to use a very simple example. Figure 2 shows the distribution of population by village in a narrow, isolated mountain valley. Given that all movement is on foot and that no services exist at present where should the basic public services be located? To simplify discussion, we assume that only sites in the villages themselves are to be considered and that for each service a constraint on resources permits only one facility.

1 Minimizing travel cost

Let p_i represent the population of any village i and n be the number of villages. The cost of travel to a service is d_{ij} , the distance from i to a facility

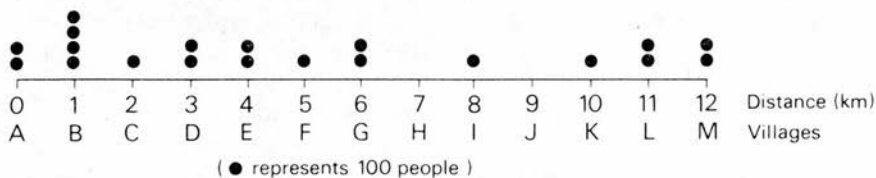


Figure 2 Distribution of population in trial example.

sited in village j ; demand or need during a certain time period is directly proportional to population and is assumed to be inelastic. If we wish to minimize the total time, cost or effort expended by the population in obtaining the service, our goal is to minimize

$$Z_A = \sum_{i=1}^n p_i d_{ij} \quad (1)$$

where Z_A is the value of this objective function at any village.

Figure 3 shows the value of aggregate travel, Z_A , at each village. This profile is completely concave, sloping upwards from the minimum at E in both directions, at first gently then more steeply.

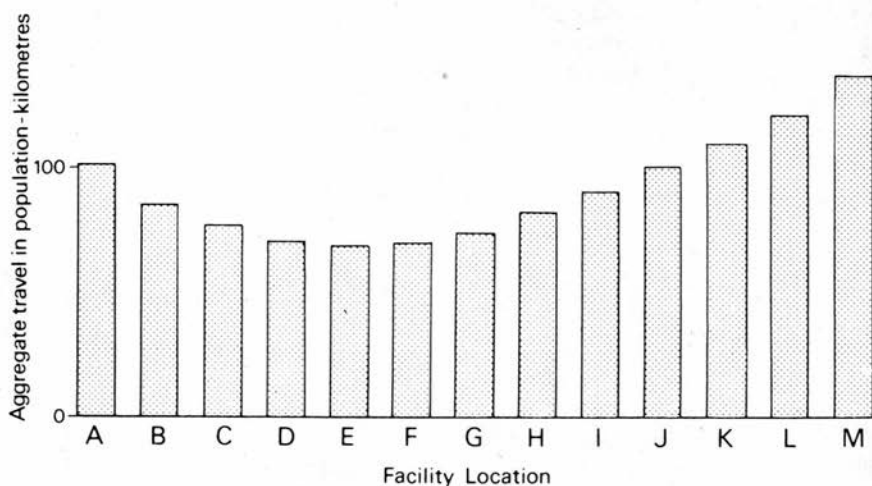


Figure 3 Aggregate travel cost at each village.

Because there is only one minimum and Figure 3 suggests that the gradient at any point runs downwards towards it, we can find the minimum quickly by a simple search process using trial and error without computing Z_A for all the villages. Simply start at any point and compute Z_A ; next compute the gradient by evaluating Z_A at a point to the left or right; then follow the direction in which the gradient is negative to the next village; continue until the gradient becomes positive; whereupon the minimum has been found. Although this account overlooks several mathematical difficulties, it can be taken as the essence of a method devised by Törnqvist (1963; 1971) to find the minimum for a distribution in two dimensions and has a broad similarity to a method using differential calculus developed by Kuhn and Kuenne (1962) and Cooper (1963).

Aggregate travel surfaces have been computed for many areas (Harris, 1954; Neft, 1966). Invariably they have one minimum and slope smoothly towards it, so that a marble placed on the surface would inevitably roll

down towards the optimum. In Figure 3 the gradient is gentle near the minimum, so that little efficiency would be sacrificed by locating at D or F or even G. Similarly, Cooper (1963, 340), Eilon *et al.* (1971), Nordbeck and Rystedt (1972) and, in slightly different context, Goodchild (1972) make the important point that the aggregate cost surface is shallow in a certain region around the minimum for very different distributions of population.

To minimize average travel we minimize

$$Z_B = \frac{\sum_{i=1}^n p_i d_{ij}}{\sum_{i=1}^n p_i} \quad (2)$$

Now, the denominator of (2),

$$\sum_{i=1}^n p_i,$$

is a constant for all locations and (2) therefore has only one variable term, namely

$$\sum_{i=1}^n p_i d_{ij}.$$

Hence, the location which minimizes aggregate travel also minimizes mean travel.

The essential properties of an aggregate travel surface are not affected by disaggregation. Suppose, for instance, that the population in each village is divided into k groups ($k=1, 2, \dots, l$) and that the average demand from any group, p_{ik} , during a certain time period is w_k . Demand from any single village is then

$$\sum_{k=1}^l w_k p_{ik}$$

so the objective becomes

$$\text{minimize } Z_C = \sum_{i=1}^n d_{ij} \sum_{k=1}^l w_k p_{ik} \quad (3)$$

If one group in the population is then known to use the service more than others it can thus be given more weight through the appropriate w_k .

Using a different argument, if a value judgement is made that one group in the population, perhaps the very old or very young, should have more influence than others on the location of the centre we can treat the w_k values simply as the importance attached to the groups, irrespective of the number of journeys each makes. In minimizing Z_C we are then no longer minimizing cost. Instead we are finding which location best balances the relative importance of the 'pulls' of different groups assuming that any point's 'pull'

increases linearly with its distance from the centre. Since the graphical and mathematical properties of the function have not essentially been changed, the methods outlined above can still be used to find the minimum.

It is well known that the point of minimum travel for a linear distribution of population is the median point on the line. The corresponding problem of locating several centres on a plane or graph is usually called the *p*-median problem, which becomes the *m*-median problem in our notation.

2 Maximizing demand

If demand is elastic and the number of trips from *i* to a facility at *j*, denoted by T_{ij} , falls at a negative exponential rate with distance from *i* then

$$T_{ij} = p_i e^{-bd_{ij}}$$

where the constant *b* describes the rate of decrease with distance.

Taylor (1971; 1975) presents a review of various ways of formulating distance decay. A negative exponential has the advantage that at zero distance from *j* the exponential expression becomes $e^{-0}(=1)$ and the demand is then p_i ; whereas a formulation of the type p_i/d^b gives infinite demand from the village which has the centre. With this formulation, the aim of locating to maximize utilization or demand can be expressed:

$$\text{maximize} \quad Z_D = \sum_{i=1}^n T_{ij} = \sum_{i=1}^n p_i e^{-bd_{ij}} \quad (4)$$

Figure 4 shows the value of Z_D at each village when $b=1$, a value which represents a relatively sharp decay, i.e. a strongly elastic demand. The profile is relatively uneven: there is a pronounced global maximum at B, a clear local maximum at L and poorly defined maxima at D, G, I. Because remote population is discounted the objective is now very sensitive to local pockets of population as is the case with population potential surfaces (Harris, 1954; Neft, 1966; Nordbeck and Rystedt, 1972). Therefore a

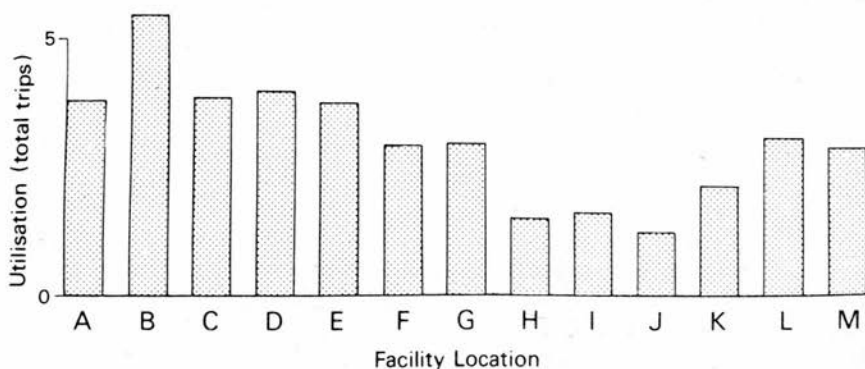


Figure 4 Total utilization of a facility at each village when $b=1$.

hill-climbing method of search which moved in the direction of ascending gradient might find only a local summit, depending on where it started. The chance of finding the global optimum could then be improved by using a series of different starting positions.

Unlike Z_A , the use maximizing function can be quite steep near the optimum, depending on how large the gradient, b , is. Since it gives good access to demand, this location might be attractive to a retailing entrepreneur (Morrill, 1974). In one sense, however, it is a very inequitable solution because it virtually ignores the demand or need of the more remote population. Hence Morrill (1974) argues that the use maximizing principle should not be used to locate essential public services; in any case it will be inappropriate if the latter have inelastic demand. However, it will be an important criterion for other services which have to raise part of their budget by charging users.

Demand maximizing models need not always be inequitable. Suppose the population in each village is divided into two groups, one of which can travel easily and has inelastic demand whereas the other finds it difficult to travel further than short distances and therefore has very elastic demand. Each group then has a different value of ' b '. In a city these groups might correspond to people with and without cars. In order to maximize demand, the solution will be relatively sensitive to the distribution of the second group but pay little heed to the more mobile group because their demand is fixed irrespective of location. A disaggregated model can therefore give more weight to the less mobile.

3 Maximizing equity

Some inequality in access is inevitable because some people will always be nearer the service node than others. To minimize this inequality, we can choose a location which reduces the longest journey of any consumer to a minimum. This 'minimax' solution is at G, 6 km from both ends of the valley. The minimax principle accords with Rawls's (1972) criterion of justice whereby the 'prospects of the least fortunate are as great as they can be'. As Morrill (1974) points out the most equitable solution on a plane with a non-uniform distribution of population is the triangular lattice of centres used in Christaller's model. The Christaller landscape on a uniform plane can thus be viewed as an extreme case in which equity and efficiency coincide perfectly. The disadvantage of a minimax solution is that it may inflict excessive travel on the majority in order to reduce travel for a few isolated users. Nevertheless it remains a significant criterion for assessing any solution. Formally the principle can be expressed as:

$$\text{minimize} \quad Z_E = \text{Max}[d_{ij}] \quad (5)$$

The centre of gravity, calculated on a plane as the mean of the weighted coordinates, minimizes the sum of the squared travel distances. It often approximates the median location and is easier to compute. Since it gives more weight to extreme distances it produces a more equitable solution than the median.

4 *Covering objectives*

A more flexible way of incorporating an element of equity into the solution is through the use of covering models (Toregas and ReVelle, 1972). For certain services, particularly emergency fire and medical services, the quality of the service or its value to the user declines with distance from the supply point. A desirable standard of service may then be defined in terms of a certain maximum time or distance, S , and the service agency may wish to position facilities to ensure that the whole population is within S units of a centre. A point within S units of a facility is then said to be covered. The fire service in Britain provides an illustration of this problem since it is recommended in urban areas that it should be able to have one pump at the scene of a fire within five minutes of the alarm being raised (Hogg, 1968).

Toregas and ReVelle (1972; 1973), have pioneered the application of covering models to problems of facility location. White and Case (1974) provide a useful review of the general family of covering models; a very clear account of the formulation and methods of solution of these problems which demonstrates their flexibility is given by Church and ReVelle (1974). An application to emergency ambulance services is discussed in ReVelle *et al.* (1976).

Suppose sufficient funds are available to provide one fire station for the valley and we wish to locate it so that as many households as possible are within the recommended distance of the centre, say 3 km. For any potential site j a binary coefficient a_{ij} can be used to describe which villages can be covered from j : when i is covered, a_{ij} is one; otherwise it is zero. The objective for this problem can then be written:

$$\text{maximize} \quad Z_F = \sum_{i=1}^n a_{ij} p_i; \quad (6)$$

$$\text{where} \quad a_{ij} = \begin{cases} 1 & \text{if } d_{ij} \leq S \\ 0 & \text{if } d_{ij} > S \end{cases}$$

$$\text{and} \quad S = 3 \text{ km.}$$

Figure 5 shows the population within 3 km cover from each village; clearly D is the best site because the whole stretch of villages from A to G can be reached from there.

The objective of the maximal covering model is more flexible and complex than the efficiency and equity models already discussed. If the covering threshold is relatively large, say 6 km, the solution is identical to the equity or minimax location, G. However, the smaller S becomes, the more the solution is attracted to the single largest pocket of population: when $S = 1$ km the best location is B, a relatively inequitable solution. Thus when there are striking contrasts in population density within a region, there will be a tension between the goals of spatial efficiency and equity. The balance between these goals can then be explored by varying the covering radius S , which could be defined equally well in terms of travel time or distance.

5 A spatial interaction model based on intervening opportunities

Some of the preceding objectives can be criticized for making the relatively simple assumption that all users go to the nearest facility. These objectives could be made somewhat more realistic by formulating the problem more explicitly in terms of spatial interaction models. To illustrate this point we can develop an intervening opportunities model for the problem.

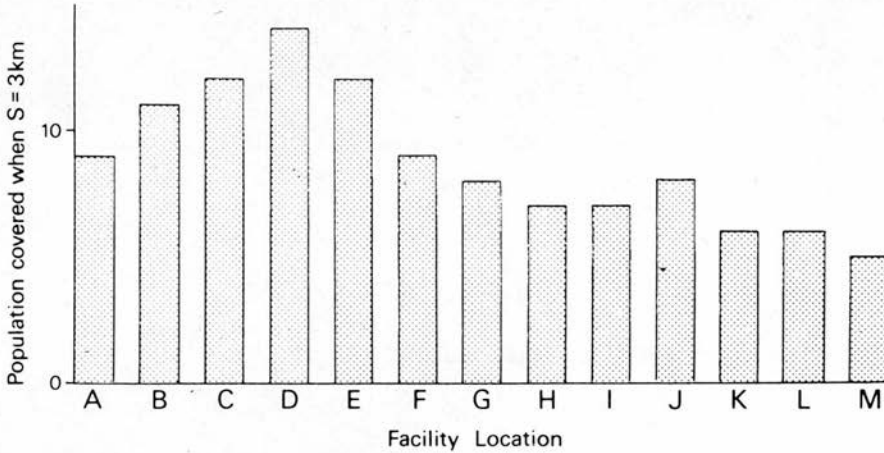


Figure 5 Population covered from a facility at each village when $S=3$.

Consider a plane or graph in which m facilities are to be located and demand is inelastic. Instead of all demand from i being satisfied by the nearest opportunity as in formulation (1), we assume that successively smaller proportions are satisfied at the first, second, third and fourth nearest facilities. Suppose q is the number of facilities closer to demand point i than is facility j , inclusive of j ; $q-1$ is the equivalent number at facility $j-1$. If the proportion of users from i who are unsatisfied, $pr(u_i)$, at any stage declines at a constant rate then

$$pr(u_i) = e^{-bq} \quad (7)$$

where b is the constant rate of decline.

The proportion satisfied at this stage will therefore be $(1 - e^{-bq})$ and the actual number $p_i(1 - e^{-bq})$. The number satisfied when the previous facility ($j-1$) had been reached would be $p_i(1 - e^{-b(q-1)})$. The number actually choosing facility j , T_{ij} , is then the difference:

$$\begin{aligned} T_{ij} &= p_i(1 - e^{-bq}) - p_i(1 - e^{-b(q-1)}) \\ &= p_i(e^{-b(q-1)} - e^{-bq}). \end{aligned} \quad (8)$$

This formulation is adapted from Zipser (1973) and is due originally to Schneider (1959).

Replacing $(e^{-b(q-1)} - e^{-bq})$ by v_j , the total number of trips from i to all j ($j=1, 2, \dots, m$) is obtained by summing over j :

$$\sum_{j=1}^m T_{ij} = \sum_{j=1}^m p_i v_j = p_i \sum_{j=1}^m v_j. \quad (9)$$

The total travel originating from i , denoted by A_i , is the product of trips and distance:

$$A_i = \sum_{j=1}^m T_{ij} d_{ij} = \sum_{j=1}^m p_i v_j d_{ij}. \quad (10)$$

Summing (9) over i the estimated total number of trips from all i in the system, denoted by T' , would be

$$T' = \sum_{i=1}^n p_i \sum_{j=1}^m v_j.$$

But since demand is inelastic the total number of trips in the system, T , is really the total demand (or population), i.e.

$$T = \sum_{i=1}^n p_i$$

This difference between the estimated total trips T' and the real total T is a standard problem in spatial interaction models and is normally resolved by scaling the number of trips from each origin to each destination up or down through multiplying by a balancing factor which consists simply of the ratio of T to T' (Wilson, 1974, 64-6). Hence the balancing factor is

$$\frac{T'}{T} = \frac{\sum_{i=1}^n p_i}{\sum_{i=1}^n p_i \sum_{j=1}^m v_j} = \frac{1}{\sum_{j=1}^m v_j} \quad (11)$$

The number of trips from i to j then becomes

$$T_{ij} = p_i v_j \left/ \sum_{j=1}^m v_j \right. \quad (12)$$

To obtain aggregate travel over the whole system, A , expression (12) is multiplied by the distance then summed over j and i :

$$A = \sum_{i=1}^n p_i \sum_{j=1}^m d_{ij} \left(v_j \left/ \sum_{j=1}^m v_j \right. \right) \quad (13)$$

The problem is then to place the facilities so that the fixed demands are satisfied with the minimal aggregate travel, i.e.

$$\text{minimize} \quad Z_G = A. \quad (14)$$

In this model catchment areas overlap, whereas in procedures for solving the p -median problem they are assumed not to overlap. The amount of overlap depends on the parameter b : disregarding the sign, a large value of b in (7) means that the unsatisfied proportion of demand falls quickly with opportunities encountered and users generally travel short distances; a small absolute value of b implies longer travel and relatively large overlapping catchment areas. As Zipser (1973) suggests, a small absolute value of b can be interpreted to mean that people are very selective and require a choice between several points of supply to meet their needs. A small absolute value of b could also reflect a generally low level of information about nearer centres.

An interesting aspect of this model is that the p -median problem can be seen as the limiting case in which all demand is satisfied by the nearest opportunity; this happens as the absolute value of b becomes very large. Though formulated in terms of intervening opportunities (or, more strictly, nearer opportunities) a very similar model could be formulated in which the proportion of unsatisfied fell directly with distance to j ; d_{ij} would then replace q as an exponent of 'e', but the model would be otherwise unchanged. Both formulations would essentially be inelastic demand models with overlapping catchment areas and a probabilistic allocation of demand. Some of the difficulties of solving such models will be discussed later. For a particular service it would be possible to estimate the crucial parameter b from data on consumer choice and travel distances. For this reason it can be claimed as a more realistic and less rigid model than the p -median model with its deterministic catchments based on Thiessen polygons.

The model just discussed could be reformulated as an elastic demand model by making the total number of trips depend on distance or travel time to the nearest facility as in Abernathy and Hershey's model (1971). The appropriate goal would then be to locate so that demand is maximized, a problem which has certain similarities to that of Lösch (1954).

Experience with spatial interaction models of a probabilistic type has not apparently been reported so far in the literature on facility location. The model has been presented here to indicate a possible avenue for future research and to illustrate the links between the two fields which have hitherto been somewhat separate, perhaps mainly because of their different historical roots: one in depot location, the other in transportation studies.

The solutions yielded by various objectives can now be compared (Figure 6). A typology of models can be created on the basis of their goals, assumptions about demand and treatment of catchment areas (Figure 7). The distinction drawn between spatial equity and efficiency is not a rigid one because all models strike a variable balance between these objectives,

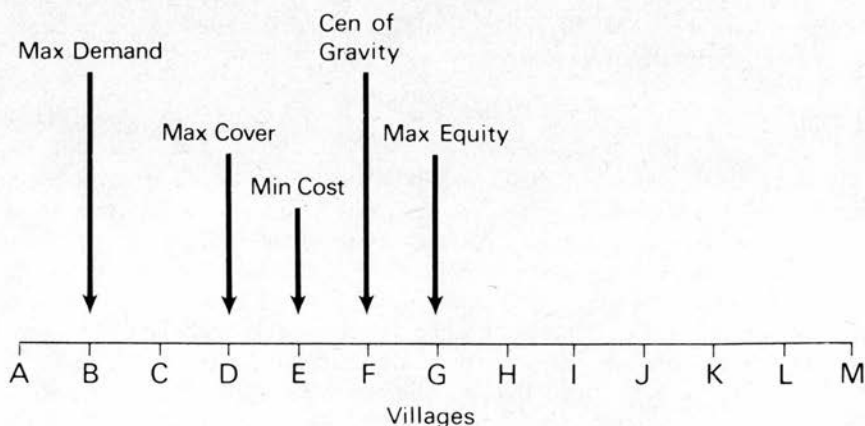


Figure 6 Optimal locations derived from different objectives.

depending on their parameters. In fact, the efficiency model for minimizing travel produced a more equitable solution in the example discussed than the equity model which maximized cover. The selection of a model to apply in a particular case is partly an empirical question about what assumptions are appropriate, but it is also a matter of social value judgments to be made politically.

IV Geometric frameworks

The choice between a network and plane is basically a choice between different sets of assumptions regarding travel by users; the geometry selected determines which methods of space searching can be used to find the optimal solution. If movement is free to take paths which approximate to straight lines or if the network of roads or railways is so dense that relatively direct paths are the rule then the main assumptions of a plane are satisfied. If these relatively demanding assumptions are not met and movement is restricted to certain routes or channels, the geometry of a graph is

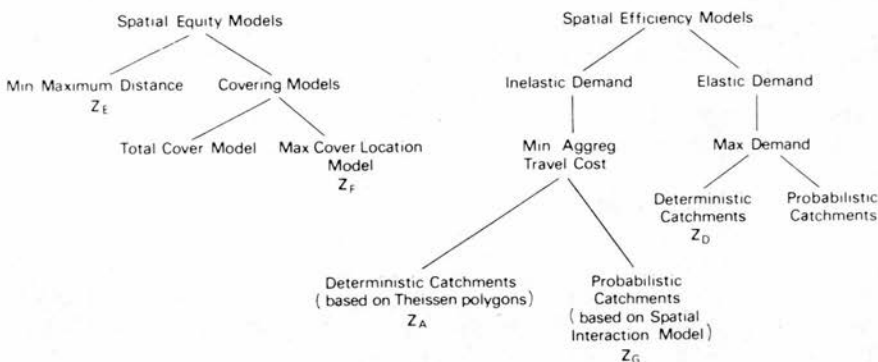


Figure 7 A classification of models for facility location.

more appropriate and travel cost can then be measured as the shortest path in time or distance through the network.

The network is clearly the more general case; in fact a plane can be viewed as a graph which has the special property that all points have straight line links to all other points. The LAP algorithm by Goodchild (Rushton *et al.*, 1973) has the useful feature of allowing distances on a plane to be measured round barriers.

The solution of the p -median problem on a graph has been made easier by a theorem proved by Hakimi (1964; 1965):

There is a set of p points, consisting entirely of nodes of the graph, which minimizes the sum of the weighted distances to the closest of any p points on the graph. (However, another set of p points, not all nodes, could possibly provide the same minimum.)

This statement of the theorem is due to ReVelle *et al.* (1970), who note that it has been extended by Levy (1967). It means that to find a single median or several medians only the nodes need to be considered, which simplifies the searching procedure. In any case, if facilities are being assigned to dispersed urban nodes within a region, the locations along the routes between the nodes will not normally be of interest because they lack the infrastructure of towns. The theorems of Hakimi and Levy seem to be among the few pieces of location theory in existence for networks.

The extra assumptions inherent in the plane purchase two advantages: first, distance can be computed directly by the Euclidean or city block metric, avoiding the need for a shortest path algorithm; second, the gradient of the objective function at any point may be obtained by using differential calculus, which facilitates search.

Strictly speaking a searching strategy should respect the intrinsic properties of the space used. Since it is not a continuous space the notion of a gradient is inappropriate for a graph. However, in the practical art of devising heuristics to obtain approximate solutions to large problems, it may not always be helpful to obey this rule too rigidly. Thus a useful procedure for examining locations on the road network of an indented coast (Robertson, 1974; 1976) makes use of Törnqvist's 'hill-climbing' strategy, thereby applying a gradient type of search to a graph. On the other hand, since a plane can be viewed as a special case of a graph, methods of searching a network could be applied on a plane by treating the points of a lattice or grid as nodes. This will usually be inefficient because it does not exploit all the information inherent in the geometry of a plane.

One apparent disadvantage of space searching procedures for a plane is that there is no way of making sure that facilities are sited in feasible locations, avoiding lakes or urban parks for instance. In practice, since the objective function for the m -median problem is shallow near the optimum, a nearby feasible location can be selected with little loss. Furthermore, these models are more likely to be used by planners to evaluate possible situations rather than to choose precise sites (Cargill and Hodgart, 1977). On a graph feasible locations can be defined as nodes so this difficulty does not arise. If the planner is only interested in evaluating a specific set of sites which are

candidates for a new facility, the Törnqvist algorithm makes it easy to do this. No searching is then done.

V Exact methods of solution

It is much easier to formulate new goals than to solve the new model efficiently. The basic difficulty is the numerical size of problems in the real world: for instance there are over six and a half million ways of assigning 3 facilities to the 342 grid cells formed when Edinburgh is divided into 500 metre grid squares. A critical test of any method for obtaining an exactly optimal or approximately optimal (i.e. heuristic) solution is the amount of computer time it requires. Much valuable information of this kind is given in an Iowa monograph on computer programs for location/allocation problems which mostly deals with *p*-median formulations (Rushton *et al.*, 1973).

1 Combinatorial programming

It is convenient to begin the discussion of exact methods by examining ways of solving the *p*-median problem on a graph and then discussing whether these methods can be used to solve models with different goals. Suppose *m* facilities are to be located to serve *n* demand nodes on a graph, how many different ways can the facilities be allocated? Since the facilities should be sited at nodes by Hakimi's theorem, this is the same as asking how many ways can *m* items be placed in *n* cells—a straightforward combinatorial problem to which the answer is

$${}_nC_m = \frac{n!}{(n-m)! m!}$$

The combinatorial character of the problem derives from the fact that any node can be given the value '1' if it has a facility and '0' if it has not. A very clear exposition of methods of optimizing a wide variety of spatial problems like this which can be formulated as a structure of zero-one variables is given in a series of studies by Scott (1969; 1971; 1975).

In order to formulate the problem in combinatorial terms a binary variable, a_{ij} , is used to describe the way demand nodes are assigned to supply points. If the demand at *i* is assigned to *j* then a_{ij} is one; otherwise, since all demand is met by the nearest facility, it is zero. Assuming there are no constraints on the capacity of facilities the objective function for a 'p-median' (i.e. an *m*-median) problem becomes:

$$\text{minimize} \quad Z_A = \sum_{i=1}^n \sum_{j=1}^n p_i d_{ij} a_{ij}. \quad (15)$$

Here a_{ij} ensures that travel from *i* to *j* is only counted when *j* is the nearest facility to *i*.

The main constraints in the problem are that only m facilities can be assigned and that all demand must be met. The whole set of constraints, derived from ReVelle and Swain (1970), can be expressed:

$$\sum_{j=1}^n a_{ij} = 1 \quad (i = 1, 2, \dots, n) \quad (15.1)$$

$$a_{jj} \geq a_{ij} \quad \begin{matrix} (i = 1, 2, \dots, n) \\ (j = 1, 2, \dots, n) \text{ and } i \neq j \end{matrix} \quad (15.2)$$

$$\sum_{i=1}^n a_{ii} = m \quad (15.3)$$

$$a_{ij} \geq 0 \quad (i = 1, 2, \dots, n) \quad (15.4)$$

$$a_{ij} = (0,1) \quad (j = 1, 2, \dots, n) \quad (15.5)$$

The first constraint ensures that all demand at i is met by ensuring that one of the a_{ij} has the value one. In models where allocation to supply points is probabilistic, a_{ij} would no longer be a zero-one variable but would be defined instead as the fraction of i 's population assigned to j . The constraint would still hold.

The second constraint ensures that the demand of a node with a facility will assign to itself rather than elsewhere, self assignment being denoted by a_{jj} or a_{ii} . When j has a facility, a_{jj} must be one, but a_{ij} may be zero or one. If there is no facility at j then both a_{jj} and a_{ij} will be zero. This constraint therefore prevents a_{ij} from having the value zero when a_{jj} is one. The third constraint restricts the number of facilities in the system to m by making use of the fact that there must be m self assigning nodes in the final solution.

Combinatorial programming involves the creation of a combinatorial tree on which each vertex describes one unique set of values for the variables to be solved, here a_{ij} . A comprehensive examination of the tree is then carried out using a systematic searching procedure such as the branch and bound or backtrack method. By this means the objective function is evaluated explicitly or implicitly at every vertex. Implicit evaluation means that it is inferred at a particular vertex that the optimal solution cannot lie on any branches descended from that point and so these branches are pruned out of the tree thereby saving computation time. The most efficient searching strategy is therefore one where a high proportion of nodes are evaluated implicitly. Unfortunately it seems to be inherently difficult in the facility location problem to eliminate any feasible set of locations by implicit evaluation, though Ostresh (1973) presents several interesting attempts to use the spatial properties of the problem on a plane to do so.

Although one can be sure of obtaining an optimal answer by these methods, their cost in computation time for large problems usually forces researchers to employ heuristic methods which obtain relatively good solutions by a rapid but incomplete search. Combinatorial methods do permit the use of non-linear objective functions since they only require the

objective function to increase or decrease monotonically as solution variables (i.e. the a_{ij}) are added. In theory they can therefore solve all the preceding formulations of the problem including the demand maximizing problem, whereas linear programming can only be used, apparently, for the p-median and covering problems (Church and ReVelle, 1974).

2 *Linear programming*

A rather surprising feature of the p-median problem, in view of its integer character, is that it can be solved fairly easily as a linear programming problem. This has been demonstrated very neatly by ReVelle and Swain (1970). Their constraint set for the problem is identical to the preceding one with the omission of the last constraint which requires a_{ij} to be one or zero. If a_{ij} is interpreted as the fraction of i 's population assigned to j all the constraints have a linear form and the problem can be solved by a standard linear programming routine.

Although a_{ij} can then by definition have any value between one and zero, any node i invariably assigns wholly to the nearest facility in order to minimize the objective function, thus yielding the desired solution of one's and zero's. The spatial structure of the problem therefore forces a binary solution onto the linear formulation. For a relatively small problem solved by ReVelle and Swain, involving six nodes and thirty facilities, only 1.51 minutes of computer time was required. The efficiency of the method for large problems has not been reported.

The dual of the linear program provides a planner with information on the marginal reduction in travel distance which would result from providing an extra facility. To solve the additional facility problem the a_{ij} values corresponding to existing facilities are given the value one.

In an extension of this model Rojeski and ReVelle (1970) replace the constraint on the number of centres by a constraint on the funds available to open new facilities or expand those already opened during the process of solution. If opening and expansion costs vary among the potential locations, this constraint has the form:

$$\sum_{j=1}^n f_j a_{jj} + \sum_{j=1}^n b_j \sum_{i=1}^n p_i a_{ij} \leq C \quad (16)$$

where f_j is the fixed cost of opening facility j

b_j is the variable cost of expanding j by one unit of population or demand

C is the investment budget.

This extended model has the advantage of allowing cost data on the supply side to be incorporated if available. Although the inclusion of an implicit constraint on capacity seems appropriate for many services (hospitals and schools for instance), it means that some customers may not be

allocated to the most convenient facility for them but rather to the cheapest one in terms of supply costs which is not already fully used. It is very unlikely that a predetermined budget will fortuitously allow an exact whole number of facilities. Let us suppose the optimal solution is nine and a half. The authors then provide the planner with a means of finding out what budgets correspond to nine and ten facilities. These alternative integer solutions can thus be compared in terms of efficiency. Though the extended model obviously has formidable data requirements, these could possibly be met in some problems where the cost of opening a new facility is being balanced against the cost of expanding an existing one.

Wagner and Falkson (1975) criticize the preceding models for assuming that demand is inelastic and argue that public services with inelastic demand are the exception rather than the rule. They then present a series of elegant models which balance the net benefits to consumers of receiving the service against the marginal cost of supplying it, thereby maximizing consumers' plus producers' surplus. Though interesting conceptually, it is very hard to envisage these models being made operational.

Because the elasticity of demand has been ignored, Falkson and Wagner argue that decisions based on fixed requirement models will result in an excessive amount of the service being supplied. In view of its significance for public policy, it is worth drawing attention to several flaws in this argument. First, fixed demand models *assume* the amount to be supplied has previously been determined in the political forum or elsewhere. It is hard to see why the supply determined in this manner should be either more or less than would result from Wagner and Falkson's model. Second, as already noted, several essential public services have inelastic demand. Third, to rely only on the pareto optimum embedded in their model is to overlook the criterion of equity which is essential to the allocation of basic public services (Olsson 1974; Dear 1974b). Furthermore, Wagner and Falkson appear to assume, in the absence of suitable evidence, that evidence of distance decay supports the assumption of elastic demand. As noted previously, distance decay rings and inelastic demand may not be inconsistent. Finally, cost minimizing models may sometimes underestimate the advantages of organizing a given supply in smaller more dispersed units (Figure 1).

VI Space searching methods for one facility on a plane

The core of many algorithms for minimizing travel on a plane is a rapid method for finding one centre, devised independently by Weiszfeld (1937), Miehle (1958), Cooper (1963) and Kuhn and Kuenne (1962). It was suggested earlier that the gradient at any point would indicate the direction of the optimum. Essentially, the method developed by Cooper and Kuhn and Kuenne is based on finding an expression for the gradient. Because of its general interest both as a method of searching space and of solving other formulations of the problem, an account of this method is given here.

We consider how the aggregate distance travelled (A_i) to a centre (X_i, Y_i)

by the population, p_i , of any point i with coordinates (x_i, y_i) changes as the centre's location moves in Euclidean space. By definition

$$A_i = p_i d_i = p_i [(x_i - X_c)^2 + (y_i - Y_c)^2]^{1/2} \quad (17)$$

As the centre moves, x_i , y_i and p_i are fixed and can therefore be treated as constants. Hence A_i can be regarded as a function of the variables X_c and Y_c , i.e.

$$A_i = f(X_c, Y_c).$$

The gradient of this function can be described by the partial derivatives with respect to X_c and Y_c . Since the partial derivative with respect to X_c describes how A_i or $f(X_c, Y_c)$ responds to an infinitely small increase in X_c when Y_c remains constant, to obtain this derivative we can treat A_i as $f(X_c)$. The partial derivative we wish to obtain is then

$$\frac{d A_i}{d X_c} = \frac{d f(X_c)}{d X_c} \quad (18)$$

To obtain the gradient we first substitute

$$u = d_i^2 = x_i^2 - 2x_i X_c + X_c^2 + y_i^2 - 2y_i Y_c + Y_c^2 \quad (19)$$

into (17). A_i , as $p_i d_i$, thus becomes $p_i u^{1/2}$. Thus A_i is a function of u . We can define this function as $g(u)$. By definition u is a function of X_c (19) which we can call h , so u can be treated as $h(X_c)$. Now since $A_i = g(u)$ and $u = h(X_c)$ aggregate travel is a function of a function of X_c .

Using the chain rule for differentiating a function of a function (Wilson and Kirby, 1975, 136)

$$\frac{d A_i}{d X_c} = \frac{d A_i}{d u} \cdot \frac{d u}{d X_c} \quad (20)$$

Of course u was chosen so that both derivatives on the right-hand side of (20) would be straightforward:

$$\frac{d A_i}{d u} = \frac{1}{2} p_i u^{-1/2} = \frac{p_i}{2 [(x_i - X_c)^2 + (y_i - Y_c)^2]^{1/2}} = \frac{p_i}{2 d_i};$$

and

$$\frac{d u}{d X_c} = -2x_i + 2X_c = -2(x_i - X_c).$$

Hence

$$\frac{d A_i}{d X_c} = \frac{p_i (-2)(x_i - X_c)}{2 d_i} = \frac{-p_i (x_i - X_c)}{d_i} \quad (21)$$

Applying the same method, the partial derivative of A_i with respect to Y_c is

$$\frac{d A_i}{d Y_c} = \frac{-p_i (y_i - Y_c)}{d_i}$$

If Z_A is the aggregate travel of all n points to the centre then

$$Z_A = \sum_{i=1}^n A_i$$

From (21) an infinitely small increase in X_c will cause each point to contribute a certain increment in travel given by the partial derivative at that point. The overall rate of change (i.e. the gradient of Z_A) will be the sum of these contributions, i.e.

$$\frac{dZ_A}{dX_c} = \sum_{i=1}^n \frac{-p_i (x_i - X_c)}{d_i}. \quad (22)$$

At the minimum point this gradient is zero by definition:

$$\sum_{i=1}^n \frac{-p_i (x_i - X_c)}{d_i} = 0. \quad (23)$$

By manipulating (23) it can be shown that

$$X_c \sum_{i=1}^n \frac{p_i}{d_i} = \sum_{i=1}^n \frac{p_i x_i}{d_i}.$$

Hence

$$X_c = \sum_{i=1}^n \frac{p_i x_i}{d_i} \bigg/ \sum_{i=1}^n \frac{p_i}{d_i}. \quad (24)$$

By symmetry

$$Y_c = \sum_{i=1}^n \frac{p_i y_i}{d_i} \bigg/ \sum_{i=1}^n \frac{p_i}{d_i}. \quad (25)$$

These expressions for the location of the minimum point cannot be solved directly because d_i is unknown. Kuhn and Kuenne found in practice, however, that it could be solved fairly easily by iterative approximation. The latter method consists of substituting arbitrary initial values of X_c and Y_c which allow the right hand side of (24) and (25) to be evaluated. This yields new values of X_c and Y_c which, as facility coordinates, have lower aggregate travel in fact than the initial values and can in turn be resubstituted back into (24) and (25) to yield a further improvement. When no further improvement is obtained the iterative procedure has converged on the minimum.

With an iterative solution there is always the possible danger that the procedure may not converge. Cases of non-convergence, however, have never been reported. In practice convergence is usually quite rapid, especially if a good initial approximation such as the centre of gravity is used. A second danger is that the minimum will be local rather than global. Since

aggregate travel surfaces for one centre are smoothly concave away from a single minimum, by descending the slope from any starting point the optimum will almost inevitably be reached. A thorough analysis of the mathematical properties and problems of the method is given by Ostresh (1973).

In contrast, the method developed by Törnqvist (1971) obtains a measure of the gradient by calculating aggregate travel to an initial trial cell and then to another cell one grid position to the west. Movement along an east-west axis then along a north-south axis continues until no improvement is recorded. The search can be carried out in successively smaller steps for greater accuracy. A complete account of this method is presented by Kohler in Rushton *et al.* (1973). If the surface of the objective function is smoothly concave away from the optimum this method will be just as successful in finding the minimum as the previous one, although it needs slightly more computation to detect the direction of the minimum. If we liken the two methods to bears searching for a barrel of honey on a pitch black night, the iterative method, like a bear with a quicker and more accurately directed sense of smell, will usually find its goal faster.

This general strategy of search outlined above could be used to optimize other objective functions. For instance, if the number of trips to a facility declines exponentially with distance so that T_i is given by $p_i e^{-bd_i}$ as in (2) then the partial derivatives for X_c and Y_c would be:

$$\frac{dT_i}{dX_c} = bp_i (x_i - X_c) e^{-bd_i} d_i^{-1}; \text{ and } \frac{dT_i}{dY_c} = bp_i (y_i - Y_c) e^{-bd_i} d_i^{-1}$$

From this it can be shown that at the point where trips are maximized

$$X_c = \frac{\sum_{i=1}^n p_i x_i e^{-bd_i} d_i^{-1}}{\sum_{i=1}^n p_i e^{-bd_i} d_i^{-1}} \quad (26)$$

and

$$Y_c = \frac{\sum_{i=1}^n p_i y_i e^{-bd_i} d_i^{-1}}{\sum_{i=1}^n p_i e^{-bd_i} d_i^{-1}} \quad (27)$$

In general the greater the absolute value of b , the more numerous become the local peaks on the surface of T_i and consequently the more likely is any searching procedure to stick on a local optimum. Computational experience with iterative solution of equations like (26) and (27) has not so far been reported in the literature, but some points where the partial derivatives are zero could be minima or saddle points.

VII Space searching heuristics for m facilities

The most efficient heuristic for solving p -median problems on a plane is the 'alternate' algorithm developed by Cooper (1963; 1967; 1968), so called

because it alternates between allocating population and locating centres in the following sequence:

- 1 assign each centre to an arbitrary initial location;
- 2 allocate each demand point to its nearest centre, defining m Dirichlet regions (*allocation stage*);
- 3 relocate each centre to the median within its catchment area by the Kuhn-Kuenne method (*location stage*);
- 4 repeat steps (2) and (3) until convergence.

Convergence must occur because steps (2) and (3) have the same objective; each can only reduce aggregate travel. The crucial difference between this procedure and the NORLOC algorithm (Törnqvist *et al.*, 1971) is that the latter only relocates *one* centre at a time in stage (3), invoking the whole allocation stage each time a single centre moves, which is quite expensive in computer time.

Both methods can obtain good solutions to the problem of locating three facilities in Edinburgh after evaluating only a few hundred of the six and a half million possible solutions, the Cooper method requiring much less time. In tests Cooper (1963) found that his method gave solutions only 2.58 per cent less efficient on average than the optimum.

If the additional facility problem is being solved, local optima become more likely because the fixed locations of existing centres constrain the search. In Figure 8 the new facility will be unable to reach the global minimum at M from a starting position at S because if it moved towards M aggregate travel would increase as it encroached on the catchments of the existing centres, fixed at A , B and C . Use of a Cooper or Törnqvist type of procedure with other objective functions will encounter the same difficulty. Repeated trials with various carefully chosen starting locations will then be needed to identify the optimum.

An 'alternate' type of algorithm for solving the p -median problem on a graph is given by Maranzana (1964). A different approach by Teitz and Bart (1968), which solves the same problem by systematic substitution of new supply nodes for those in solution, finds the optimum more frequently but requires more computer time (Rushton *et al.*, 1973). In future it may be

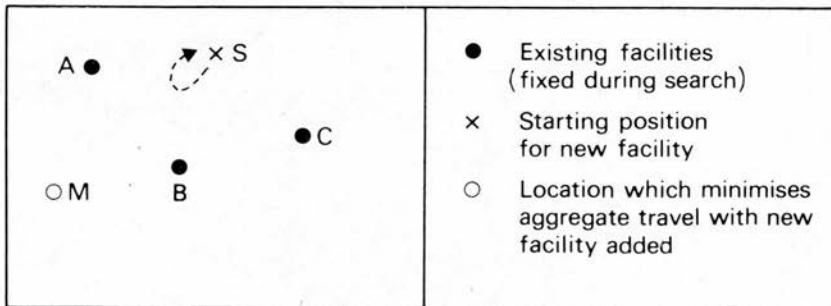


Figure 8 Search restricted by existing facilities.

possible to treat problems on a more complex, but more realistic, network structure which integrates travel on bus and rail sub-networks and treats car journeys separately.

One way of using these algorithms is illustrated by Figure 9 which shows the actual locations of public swimming pools in Edinburgh in 1971 and the optimal locations computed by NORLOC. Aggregate travel to the present pools is 42 per cent greater than to the optimal locations. In view of this it is surprising that the only pool constructed in the city since 1939, built for the Commonwealth Games in 1970, was sited in an area relatively well provided already (B) rather than on the western side of the city where there are still no public pools. This location was probably chosen because of its short-term advantage of being near the stadium (A) and athletes' accommodation used for the games, but it is hard to understand why spatial efficiency and equity were, apparently, neglected.

An 'alternate' type of algorithm could be written for other models including the demand maximizing formulation and the model with overlapping hinterlands. In the former case the location stage could find the point of maximum use by solving equations (26) and (27) iteratively or by a Törnqvist type of search. In the latter case the location stage would minimize aggregate travel within each of the overlapping catchments, whereas the allocation stage would simply redistribute demand probabilistically, which might increase aggregate travel. Since the two stages would not be pursuing the same objective, the algorithm might not converge sometimes. Clearly experiment is needed with these approaches; this would seem to be a useful area for future research.

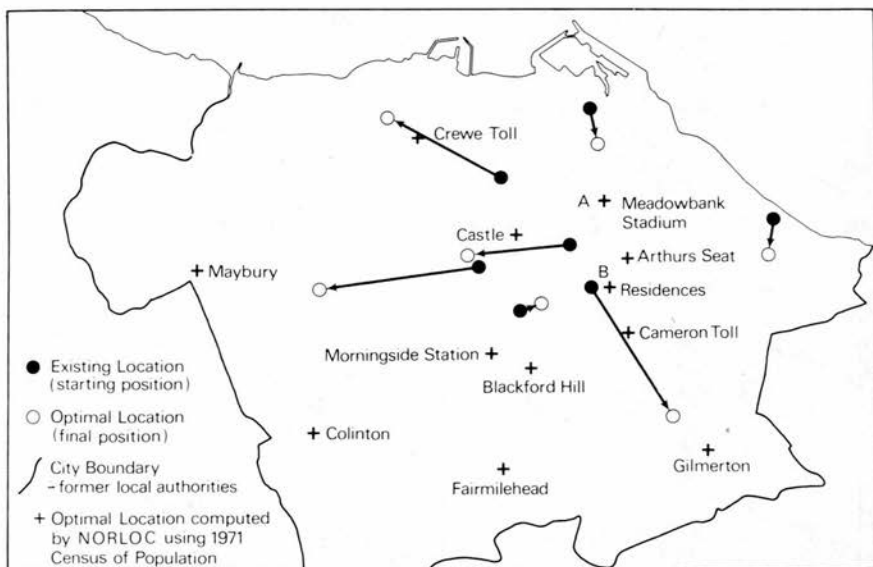


Figure 9 Comparison of actual locations of public swimming pools in Edinburgh in 1971 with locations which minimize aggregate travel.

VIII Constraints on capacity

In the models discussed so far, capacity has been entirely dependent on location; this will tend to yield small facilities in areas of low density and large ones in areas with high density of population. If users are free to choose facilities and if capacity does not normally affect the user's access to the service (e.g. libraries and, most of the time, swimming pools), this dependence seems quite acceptable.

On the other hand when schools or hospitals are involved (Yeates, 1963; Gould and Leinbach, 1966) the capacities of existing centres should be allowed to influence the locations, size and tributary areas of any additional facilities. This can be done by specifying capacities which remain fixed throughout the process of solution (Garrison, 1959) as in the LAP algorithm used by Goodchild and Massam (1969) to study administrative areas. The allocation stage then involves solving a linear programming transportation problem defined by the locations at that stage of search and by the given capacities (Cooper, 1972).

These capacities could reflect factors on the supply side such as economies of scale or the greater attractiveness of large facilities (ReVelle and Church, 1976). Many public services, however, are labour intensive and consequently have insignificant economies of scale (Hirsch, 1968). A simpler approach may be to add to the 'unconstrained' model a constraint which requires facilities to be greater than a minimal size needed for efficient operation. A different way of treating capacities has been developed by Öberg (1976) who measured variations in access to dental services in Sweden by the distance an individual had to travel to find a dentist not fully utilized by nearer patients.

IX Conclusion

It is to be hoped that teachers and researchers in geography and planning will show greater awareness of the different ways of formulating and solving multiple location problems and of the attendant assumptions and value judgements involved. The new methods, nevertheless, extend traditional graphical methods rather than replace them. Armed only with detailed maps of the relevant population groups, many individuals can identify efficient solutions to some kinds of facility location problem as Massam (1975, 94-5) points out. The advantage of computer algorithms is that they can easily evaluate locations with respect to a variety of goals and explore the effect of giving different weights to social groups according to their need, demand and mobility. The advantages of both approaches may eventually be combined through interactive computer graphics.

Improving geographical access to public services is only part of a much wider problem. Webber (1973) has written of the significant cultural and social barriers between working-class users and the professionals who administer educational, medical and other services. Furthermore it has been

shown that distance is more of a barrier to less affluent and other disadvantaged groups (Weiss and Greenlick, 1970; Forster, 1972). Hence for many services the social and spatial dimensions of the problem should be considered together. Such an approach no longer presents intractable problems but its possibilities are still to be fully realized.

University of Edinburgh

Acknowledgements

I am very grateful to Gordon Ewing of McGill University and to John Martin and Ken McKinnon of Edinburgh University for their helpful comments on my work. I would like to thank Ray Harris for the diagrams. I would also like to thank Peter Gould (Pennsylvania State University) and Elma Robertson (University of Glasgow) for their help and encouragement at an early stage in my work.

X References

- Abernathy, W. J. and Hershey, J. C.** 1971: A spatial allocation model for regional health service planning. *Operations Research* 19, 629-42.
- Abler, R., Adams, J. S. and Gould, P.** 1971: *Spatial organisation*. New Jersey: Prentice Hall. (587 pp.)
- Austin, C. M., Smith, T. E. and Wolpert, J.** 1970: The implementation of controversial facility-complex programs. *Geographical Analysis* 2, 315-29.
- Austin, C. M.** 1974: The evaluation of urban public facility locations: an alternative to benefit-cost analysis. *Geographical Analysis* 6, 135-45.
- Berry, B. J. L.** 1967: *Geography of market centers and retail distribution*. New Jersey: Prentice Hall. (146 pp.)
- Bradley, J. E., Kirby, A. M. and Taylor, P. J.** 1976: *Distance decay and dental decay*. University of Newcastle upon Tyne, Department of Geography, Seminar Paper No. 31. (23 pp.).
- Bunge, W.** 1962: *Theoretical geography*, first edition. *Lund Studies in Geography, Series C, General and Mathematical Geography* 1, Lund: Gleerup. (210 pp.)
- Cargill, S. and Hodgart, R. L.** 1977: *A strategy for the provision of swimming pools in the Lothian Region: an application of locational optimising model*. Department of Geography, University of Edinburgh. Research report prepared for Lothian Regional Council (30 pp.)
- Church, R. and ReVelle, C.** 1974: The maximal covering location problem. *Papers of the Regional Science Association* 32, 101-18.
- Church, C. L. and ReVelle, C. S.** 1976: Theoretical and computational links between the p-median, location set covering and the maximal covering location problem. *Geographical Analysis* 8, 406-15.
- Coates, B. E. and Rawstron, E. M.** 1971: *Regional variations in Britain: studies in economic and social geography*. London: Batsford. (304 pp.)
- Cooper, L.** 1963: Location-allocation problems. *Operations Research* 11, 331-43.
- Cooper, L.** 1967: Solutions of generalised locational equilibrium models. *Journal of Regional Science* 7, 1-18.

- Cooper, L.** 1968: An extension of the generalised Weber problem. *Journal of Regional Science* 8, 181-98.
- Cooper, L.** 1972: The transportation location problem. *Operations Research* 20, 94-108.
- Coppock, J. T. and Duffield, B. S.** 1975: *Recreation in the countryside*. London: Macmillan. (262 pp.)
- Currie, I. M.** 1977: *Edinburgh's indoor swimming baths: location, use and catchment areas*. University of Edinburgh, Department of Geography, M.A. thesis. (70 pp.)
- Davies, B. P.** 1968: *Social needs and resources in local services*. London: Joseph. (334 pp.)
- Dear, M. J.** 1974a: A paradigm for public facility location theory. *Antipode* 6, 46-50.
- Dear, M.** 1974b: *A political theory of public facility location*. Paper presented at the 21st North American Meetings of the Regional Science Association, Chicago.
- de Vise, P.** 1973: *Misused and misplaced hospitals and doctors*. Washington, DC: Association of American Geographers, Commission on College Geography, Resource Paper 22.
- Doherty, J.** 1973: *The location of health facilities in Tanzania*. Bureau of Resource Assessment and Land Use Planning, University of Dar es Salaam, Service Paper 73/2. (10 pp.)
- Eilon, S., Watson-Gandy, C. D. T. and Christofides, N.** 1971: *Distribution management: mathematical modelling and practical analysis*. London: Griffin. (240 pp.)
- Eilon, S.** 1972: Goals and constraints in decision making. *Operational Research Quarterly* 23, 3-15.
- Forster, F.** 1972: *The geographic viewpoint in medical care studies*. University of Edinburgh, Department of Community Medicine, PhD thesis.
- Galbraith, J. K.** 1974: *Economics and the public purpose*. London: Deutsch. (334 pp.)
- Garrison, W. L.** 1959: Spatial structure of the economy, II. *Annals of the Association of American Geographers* 49, 471-82.
- Goodchild, M. F. and Massam, B. H.** 1969: Some least cost models of spatial administrative systems in southern Ontario. *Geografiska Annaler* 52, 86-94.
- Goodchild, M. F.** 1972: The trade area of a displaced hexagonal lattice point. *Geographical Analysis* 4, 105-7.
- Gould, P. R. and Leinbach, T. R.** 1966: An approach to the geographic assignment of hospital services. *Tijdschrift Voor Economische en Sociale Geografie* 57, 203-6.
- Hakimi, S.** 1964: Optimum locations of switching centers and the absolute centers and medians of a graph. *Operations Research* 12, 450-59.
- Hakimi, S.** 1965: Optimum distribution of switching centers in a communications network and some related graph-theoretic problems. *Operations Research* 13, 462-75.
- Harris, C. D.** 1954: The market as a factor in the localization of industry in the United States. *Annals of the Association of American Geographers* 44, 315-48.
- Harvey, D.** 1973: *Social justice and the city*. London: Edward Arnold. (336 pp.)
- Hirsch, W. Z.** 1968: The supply of urban public services. In Perloff, H. and Wingo L., editors, *Issues in urban economics*, Baltimore: The Johns Hopkins University Press, 477-525.

- Hogg, J.** 1968: The siting of fire stations. *Operational Research Quarterly* 19, 275-87.
- Hotelling, H.** 1929: Stability in competition. *Economic Journal* 41, 41-57.
- Hume, D.** 1930: *An enquiry concerning the principles of morals*. Chicago: Open Court Publishing Co. (169 pp.)
- Kuehn, A. and Hamburger, N.** 1963: A heuristic program for locating warehouses. *Management Science* 9, 643-66.
- Kuhn, H. W. and Kuenne, R. E.** 1962: An efficient algorithm for the numerical solution of the generalized Weber problem in spatial economics. *Journal of Regional Science* 4(2), 21-34.
- Lea, A. C.** 1973: *Location—allocation systems: an annotated bibliography*. University of Toronto, Department of Geography, Discussion Paper 13. (273 pp.)
- Levy, J.** 1967: An extended theorem for location on a network. *Operational Research Quarterly* 18, 535-42.
- Lösch, A.** 1954: *The economics of location*. New Haven: Yale University Press. (520 pp.)
- McAllister, D. M.** 1976: Equity and efficiency in public facility location. *Geographical Analysis* 8, 47-63.
- McGrew, J. C. and Monroe, C. B.** 1975: Efficiency, equity and multiple facility location. *Proceedings of the Association of American Geographers* 7, 142-6.
- Maranzana, F. E.** 1964: On the location of supply points to minimise transport costs. *Operational Research Quarterly* 15, 261-70.
- Massam, B. H.** 1972: *The spatial structure of administrative systems*. Washington, DC: Association of American Geographers, Commission on College Geography, Resource paper 12.
- Massam, B.** 1974: Political geography and the provision of public services. *Progress in Geography* 6, 179-210.
- Massam, B.** 1975: *Location and space in social administration*. London: Arnold. (192 pp.)
- Miehle, W.** 1958: Link length minimization in networks. *Operations Research* 6, 232-43.
- Morrill, R. L.** 1974: Efficiency and equity of optimum location models. *Antipode* 6, 41-6.
- Mumhrey, A. J. and Wolpert, J.** 1973: Equity considerations and concessions in the sitings of public facilities. *Economic Geography* 44, 109-21.
- Navarro, V.** 1976: Social class, political power and the state and their implications in medicine. *Social Science and Medicine* 10(9/10), 437-57.
- Neft, D. S.** 1966: *Statistical analysis for areal distributions*. Regional Science Research Institute, Monograph Series 2.
- Nordbeck, S. and Rystedt, B.** 1972: *Computer cartography*. Lund: Studentlitteratur. (315 pp.)
- Öberg, S.** 1976: *Methods of describing physical access to supply points*. Lund Studies in Geography, Series B, Human Geography 43, Lund: Gleerup. (141 pp.)
- Olsson, G.** 1974: Servitude and inequality in spatial planning: ideology and methodology in conflict. *Antipode* 6, 16-21.
- Ostresh, L. M.** 1973: *An investigation of the multiple location—allocation problem*. University of Iowa, Department of Geography, PhD thesis. (213 pp.)
- Pahl, R. E.** 1975: *Whose city?* London: Penguin. (306 pp.)
- Polarczyk, K.** 1976: The distribution of service centres within large urban areas. A market accessibility model. *Geographia Polonica* 33, 143-55.

- Rawls, J.** 1972: *A theory of justice*. Oxford: Oxford University Press. (607 pp.)
- ReVelle, C. S. and Swain, R. W.** 1970: Central facilities location. *Geographical Analysis* 2(1), 30-42.
- ReVelle, C. Marks, D. and Liebman, J. C.** 1970: An analysis of private and public sector location models. *Management Science* 16, 692-707.
- ReVelle, C. and Church, R.** 1976: *A spatial model for the location construct of Teitz*. Paper presented at Toronto meeting of the Regional Science Association. (10 pp.)
- ReVelle, C., Toregas, C. and Falkson, L.** 1976: Applications of the location set-covering problem. *Geographical Analysis* 8, 65-76.
- Robertson, I. M. L.** 1974: Road networks and the location of facilities. *Environment and Planning A* 6, 199-206.
- Robertson, I. M. L.** 1976: Accessibility to services in the Argyll District of Strathclyde: a locational model. *Regional Studies* 10, 89-95.
- Rojeski, P. and ReVelle, C.** 1970: Central facilities location under an investment constraint. *Geographical Analysis* 2, 343-60.
- Rushton, G., Goodchild, M. F. and Ostresh, L. M.** 1973: *Computer programs for location/allocation problems*. Iowa: Department of Geography, University of Iowa. (321 pp.)
- Schneider, M.** 1959: Gravity models and trip distribution theory. *Papers of the Regional Science Association* 5, 51-6.
- Scott, A. J.** 1969: On the optimal partitioning of spatially distributed point sets. In Scott, A. J., editor, *Studies in regional science*, London: Pion.
- Scott, A. J.** 1970: Location-allocation systems: a review. *Geographical Analysis* 2, 95-119.
- Scott, A. J.** 1971: *Combinatorial programming, spatial analysis and planning*. London: Methuen. (204 pp.)
- Scott, A. J.** 1975: Discrete dynamic locational systems: notes on a structural framework. In Karlqvist, A., Lundqvist, L. and Snickars, F., editors, *Dynamic allocation of urban space*, Farnborough, England: Saxon House. (383 pp.)
- Smolensky, E., Burton, R. and Tideman, N.** 1970: The efficient provision of a local non-private good. *Geographical Analysis* 2, 330-42.
- Taylor, P. J.** 1975: *Distance decay in spatial interactions*. IBG Study Group in Quantitative Methods, CATMOG 2. Norwich: Geo Abstracts.
- Taylor, P. J.** 1971: Distance transformation and distance decay functions. *Geographical Analysis* 3, 221-38.
- Taylor, P. J.** 1977: *Quantitative methods in geography*. Boston: Houghton Mifflin. (386 pp.)
- Teitz, M. B.** 1968: Toward a theory of urban public facility location. *Papers of the Regional Science Association* 21, 35-51.
- Teitz, M. B. and Bart, P.** 1968: Heuristic methods for estimating the generalized vertex median of a weighted graph. *Operations Research* 16, 901-1092.
- Toregas, C. and ReVelle, C.** 1972: Optimal location under time or distance constraints. *Papers of the Regional Science Association* 28, 133-43.
- Törnqvist, G.** 1963: *Studier i industrilokalisering*. Stockholm: Geografiska Institutionen vid Stockholms universitet.
- Törnqvist, G., Nordbeck, S., Rystedt, B. and Gould, P.** 1971: *Multiple location analysis*. Lund Studies in Geography, Series C, General and Mathematical Geography 12, Lund: Gleerup.

- Wagner, J. L. and Falkson, L. M.** 1975: The optimal nodal location of public facilities with price sensitive demand. *Geographical analysis* 7, 69–83.
- Webber, M.** 1973: Planning in an environment of change. In Cullingworth, J. B., editor, *Problems of an urban society*, volume 3, London: Allen and Unwin. (195 pp.)
- Węclawowicz, G.** 1975: *Struktura przestrzeni społeczno-gospodarczej Warszawy w latach 1931 i 1970 w swietle analizy czynnikowej*. (The structure of socio-economic space of Warsaw in 1931 and 1970 in the light of factor analysis.) Polska Akademia Nauk Instytut Geografii i Przestrzennego Zagospodarowania, Prace Geograficzne 116, Warsaw.
- Weiss, J. E. and Greenlick, M. R.** 1970: Determinants of medical care utilization: the effect of social class and distance on contacts with the medical care system. *Medical Care* 8, 456–62.
- Weiszfeld, E.** 1937: Sur le point pour lequel la somme des distances de n points donnés est minimum. *Tohuko Math Journal* 43, 355–86.
- White, J. A. and Case, K. E.** 1974: On covering problems and the central facilities location problem. *Geographical Analysis* 6, 281–93.
- Wilson, A.** 1974: *Urban and regional models in geography and planning*. London: Wiley. (418 pp.)
- Wilson, A. G. and Kirkby, M. J.** 1975: *Mathematics for geographers and planners*. Oxford: Oxford University Press. (325 pp.)
- Wolpert, J., Mumphrey, A. and Seley, J.** 1972: *Metropolitan neighborhoods: participation and conflict over change*. Washington, DC: Association of American Geographers, Commission on College Geography, Resource Paper 16.
- Wolpert, J., Dear, M. and Crawford, R.** 1975: Satellite mental health facilities. *Annals of the American Association of Geographers* 65, 24–35.
- Yeates, M.** 1963: Hinterland delimitation: a distance minimising approach. *Professional Geographer* 15(6), 7–10.
- Zipser, T.** 1973: A simulation model of urban growth based on the model of the opportunity selection process. *Geographia Polonica* 27, 119–32.